

MASTER'S THESIS

Implementatie van Duurzaam Beoordelen.

Een GCM Studie voor Praktijk- en Theorie-Geïnformeerde Ontwerpprincipes voor het Middelbare Beroepsonderwijs.

Keetels, Sanne

Award date:
2021

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

Open Universiteit
www.ou.nl





Implementatie van Duurzaam Beoordelen:

Een GCM Studie voor Praktijk- en Theorie-Geïnformeerde Ontwerpprincipes voor het Middelbare
Beroepsonderwijs.

Implementation of Sustainable Assessment:

A GCM Study for Practice and Theory Informed Design Principles for Vocational Education.

Drs. Sanne Keetels

Master Onderwijswetenschappen
Open Universiteit

Cursusnaam en cursuscode:	Masterthesis OM9906182214M
Naam begeleider:	Dr. M. R. van Diggelen
Datum:	19-04-2021

Inhoud

Samenvatting.....	3
Summary.....	4
1. Inleiding.....	5
1.1 Theoretische kader.....	6
1.1.1 Beoordelen voor leren en ontwikkelen: formatief en duurzaam.....	6
1.1.2 Opvattingen van docenten.....	11
1.2 Vraagstellingen en hypothesen.....	12
2. Methode.....	13
2.1 Ontwerp.....	13
2.2 Participanten.....	14
2.3 Materialen.....	15
2.4 Procedure.....	15
2.5 Data-analyse.....	17
3. Resultaten.....	19
3.1 Resultaten sorteerddata.....	19
3.2 Resultaten scoringsdata.....	24
4. Discussie en conclusie.....	28
Referenties.....	35
Bijlagen.....	38

**Implementatie van Duurzaam Beoordelen:
Een GCM Studie voor Praktijk- en Theorie-Geïnformeerde Ontwerpprincipes voor het
Middelbare Beroepsonderwijs.**

Samenvatting

Duurzaam beoordelen biedt een nieuw perspectief om de bestaande beoordelingspraktijk in het mbo te hervormen naar een praktijk die bijdraagt aan het leven lang leren van een student. Het implementatieproces is echter moeizaam en vindt nog onvoldoende plaats, omdat de opvattingen van docenten over duurzaam beoordelen in de huidige wetenschappelijke literatuur ontbreekt, terwijl zij de centrale schakel zijn voor een succesvolle implementatie. Het doel van dit ontwerpgerichte onderzoek is om breedgedragen ontwerpprincipes te formuleren ter bevordering van de duurzame beoordelingspraktijk van de participerende mbo-instelling, waarin zowel opvattingen van docenten als wetenschappelijke inzichten elkaar complementeren. Hiertoe is het mixed method ontwerp van group concept mapping (GCM) gebruikt. Er hebben 27 mbo-docenten deelgenomen aan het onderzoek. Door middel van een brainstorm zijn 73 stellingen over duurzaam beoordelen verzameld door 27 docenten. Deze stellingen zijn aangevuld met 11 ontwerpuitgangspunten uit de wetenschappelijke literatuur. De totale statements-set bestaande uit 84 stellingen, is door 19 docenten gesorteerd en door 11 docenten gescoord op belang en haalbaarheid. Een multi-dimensionele schaalanalyse en hiërarchische clusteranalyse resulteerden in een 8-cluster concept map. De clusters waren: 1) *Voorwaarden voor beoordelingssituaties*, 2) *Docent als coach*, 3) *Hulpmiddelen*, 4) *Zelfsturing*, 5) *Organisatiecultuur*, 6) *Voorwaarden voor implementatie*, 7) *Ipsatief beoordelen*, 8) *Groeimindset*. Uit de resulterende conceptmap met cluster-indelingen en cluster-scores, zijn ontwerpprincipes afgeleid die een gerede kans hebben door docenten in de praktijk te worden geïmplementeerd. Daarnaast zijn ook de in de theorie beschreven uitgangspunten van duurzaam beoordelen vanuit het praktijkperspectief van docenten gevalideerd en gecontextualiseerd.

Keywords: leven lang leren, duurzaam beoordelen, opvattingen docenten, ontwerpprincipes, mbo, group concept mapping

Summary

Sustainable assessment offers new insights into the assessment practices in vocational education by concentrating on the future learning of learners, thus contributing to their lifelong learning skills. However, the implementation process is challenging because the scientific literature on sustainable assessment is predominantly conceptual and does not include teachers' beliefs, even though they play an essential part in educational innovations. This design based research aims to formulate design principles that incorporate both scientific and teacher's beliefs on sustainable assessment, to further the sustainable assessment practice of the participating educational institution. To this end, the mixed method design of group concept mapping (GCM) has been employed. A total of 27 vocational-teachers participated in this research. During a brainstorm session with these 27 participants, 73 statements on sustainable assessment were generated. Eleven literature based statements were added to this statement-set, resulting in a total of 84 unique statements. These 84 statements were sorted by 19 participants and scored on importance and attainability by 11 participants. Multidimensional scaling and hierarchical cluster analysis resulted in a 8-cluster concept map. The cluster labels were: 1) *Conditions for assessment situations*, 2) *Teacher as coach*, 3) *Tools*, 4) *Self-management*, 5) *Organisational culture*, 6) *Conditions for implementation*, 7) *Ipsative assessment*, and 8) *Growth mindset*. From these clusters, eight design principles were formulated, which might help teachers further their sustainable assessment practice. Furthermore, the results help to validate and contextualise the literature based insights from a practical perspective.

Keywords: lifelong learning, sustainable assessment, teachers' beliefs, design principles, vocational education, group concept mapping.

1. Inleiding

Mbo-instellingen moeten hun studenten voorbereiden op een leven lang leren, omdat kennisgebieden, werkwijzen en technieken in de huidige kenniseconomie voortdurend veranderen (Onderwijsraad, 2016). Dit vraagt naast andere onderwijsdoelen ook om andere manieren van begeleiden en beoordelen. De certificerende functie van beoordelen (summatief) is niet meer toereikend om te bepalen of studenten voldoende voorbereid zijn op de toekomst. Het is in toenemende mate belangrijk om de student te voorzien van rijke informatie die verder leren bevordert (formatieve functie). Daarnaast is het belangrijk dat beoordelen gericht is op de individuele ontwikkeling van de student met het oog op zijn of haar toekomst na de initiële opleiding, de zogenaamde duurzame functie van beoordelen (Boud, 2000; Sluijsmans, 2010).

Het realiseren van een duurzame beoordelingspraktijk is echter niet eenvoudig en vindt nog onvoldoende plaats (Bennett, 2011; Gulikers & Baartman, 2017; Hattie & Timperley, 2007; Schildkamp et al., 2014; Sluijsmans, Joosten-ten Brinke, & Van der Vleuten, 2013). Het moeizame implementatieproces is onder andere te wijten aan de conceptuele aard van de bestaande wetenschappelijke literatuur omtrent duurzaam beoordelen, wat een vertaalslag naar de praktijk in de weg staat (Boud & Soler, 2016; Fastré, Van der Klink, Sluijsmans, & Merriënboer, 2012). Een andere oorzaak is te herleiden naar het implementatieproces zelf, waarbij van docenten verwacht wordt dat zij de in de literatuur beschreven uitgangspunten van duurzaam beoordelen vertalen naar de dagelijkse praktijk. Hierbij wordt vaak onvoldoende rekening gehouden met de opvattingen van de docenten zelf, terwijl zij de centrale schakel zijn in het bewerkstelligen van onderwijsvernieuwingen (Gulikers & Baartman, 2017; Sluijsmans, 2010; Thurlings & Van Diggelen, 2019; Verloop, Van Driel, & Meijer, 2001). Opvattingen van docenten over beoordelen beïnvloeden sterk de invulling van de beoordelingspraktijk en zijn daarom voorwaardelijk om onderwijsinnovaties te laten slagen (Calderhead, 1996; Pajares, 1992; Sadler, 1989; Thomas, 2012; Verloop et al., 2001). Wanneer de opvattingen van docenten niet in overeenstemming zijn met de wetenschappelijke literatuur, is het waarschijnlijk dat de eigen bestaande opvattingen de beoordelingspraktijk bepalen (Van Berkel, Bax, & Joosten-ten Brinke, 2017; Sluijsmans et al., 2013). Tegen deze achtergrond heeft een mbo-instelling in het Zuiden van Nederland de ambitie een ontwerpgericht onderzoek uit te voeren om een duurzame beoordelingspraktijk te realiseren. Het voorliggende voorstel is onderdeel van de eerste fase van het ontwerpgericht onderzoek.

Om de vertaalslag van theorie naar praktijk te bevorderen heeft dit ontwerpgerichte onderzoek het doel om ontwerpprincipes te formuleren, waarin zowel docenten opvattingen als wetenschappelijke inzichten elkaar complementeren. Het nadrukkelijk betrekken van docenten bij vormgeven van innovaties creëert bovendien draagvlak en door aan te sluiten en voor te bouwen op de opvattingen van docenten wordt een duurzamer resultaat bereikt. Om die reden wordt group concept mapping GCM ingezet als onderzoeksmethodologie, omdat het de mogelijkheid biedt tot co-creatie. De

opvattingen van docenten die het meest met het onderwerp van doen hebben (group wisdom) kunnen met GCM in dialoog worden gebracht met de in de literatuur beschreven uitgangspunten. De cluster-indelingen en itemscores die door middel van GCM gegenereerd worden, bieden een basis waaruit ontwerpprincipes kunnen worden geformuleerd die een gereede kans hebben door docenten in de praktijk te worden geïmplementeerd. Op die manier levert dit onderzoek theoretische- en docent-geïnformeerde ontwerpprincipes op ter bevordering van een duurzame beoordelingspraktijk binnen de participerende onderwijsinstelling en worden de in de theorie beschreven uitgangspunten van duurzaam beoordelen vanuit een praktijkperspectief gevalideerd en gecontextualiseerd.

1.1 Theoretische kader

In dit onderzoek staan de concepten duurzaam beoordelen en opvattingen van docenten centraal. Beide concepten worden besproken in dit theoretisch kader, dat eindigt met een overzicht van relevante ontwerpprincipes over duurzaam beoordelen afgeleid uit de wetenschappelijke literatuur.

1.1.1 Beoordelen voor leren en ontwikkelen: formatief en duurzaam

Om een leven lang te blijven leren, moet de student leren zijn eigen prestaties te beoordelen, om blijvend en zelfstandig richting te kunnen geven aan zijn ontwikkeling (Boud, 2000; Sluijsmans, 2010). Om die reden is er zowel maatschappelijk als wetenschappelijk toenemende aandacht voor andere functies van beoordelen, die het leerproces ondersteunen (Van Berkel et al., 2017). De formatieve functie van beoordelen, waarin de focus ligt op het verbeteren van de leerprestaties, kan hieraan bijdragen (Fastré et al., 2012). Bij formatieve beoordelingen gaat het om het dichten van de kloof tussen de huidige prestatie en de gewenste prestatie door middel van het proces van zoeken, aggregeren en interpreteren van informatie die studenten en docenten gebruiken bij het bepalen waar studenten staan in hun leerproces, waar zij naartoe moeten werken en op welke manier (Assessment Reform Group, 2002; Sluijsmans et al., 2013; Van Berkel et al., 2017).

De formatieve functie van beoordelen is een goede stap in de richting van de doelstellingen van een leven lang leren, maar Black en Wiliam (1998) concluderen in hun reviewstudie dat in de praktijk nog steeds onvoldoende nadruk wordt gelegd op de zelfsturende vaardigheden van de student. Volgens Black en Wiliam wordt er in de praktijk nauwelijks onderscheid gemaakt tussen het geven van feedback en beoordelen en wordt er weinig gebruik gemaakt van formatieve methoden zoals self- en peerassessment, terwijl juist deze methoden geschikt zijn om de zelfsturende vaardigheden van de student te bevorderen (Black & William, 1998). Ook andere auteurs zoals Boud (2000) en Sluijsmans (2008) bevestigen dat met formatief beoordelen de stap naar meer zelfsturing van de student maar moeizaam gemaakt wordt. Daarom introduceert Boud (2000) het concept duurzaam beoordelen als een derde functie van beoordelen. Bij het uitwerken van het concept duurzaam beoordelen bouwt Boud voort op literatuur over formatief beoordelen, in het bijzonder op het werk van Black en William (1998). Het grote verschil dat Boud aanbrengt ten opzichte van formatief beoordelen is dat duurzaam

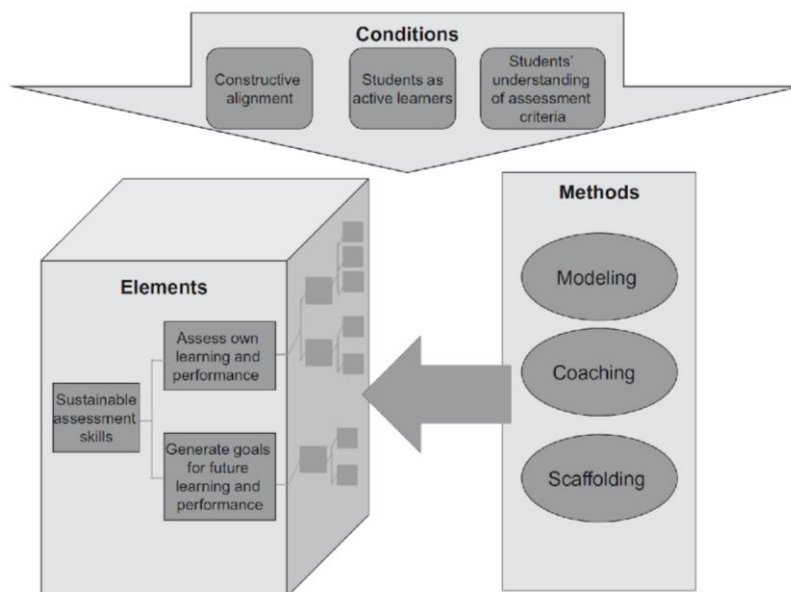
beoordelen niet alleen beoogt het leren binnen de opleiding te bevorderen, maar ook het toekomstige leren buiten de opleiding (Boud, 2000). Het belangrijkste doel van duurzaam beoordelen volgens Boud, is het aanleren van zelfbeoordelingsvaardigheden bij studenten, zodat zij ook na de opleiding verantwoordelijkheid voor hun eigen professionele ontwikkeling kunnen nemen.

Het concept duurzaam beoordelen is door Boud en collega's in de jaren daarna verder ontwikkeld tot een theoretisch raamwerk bestaande uit vier conceptuele kenmerken (Boud, 2010). Het eerste kenmerk, wat kan worden gezien als het doel van duurzaam beoordelen, is dat beoordelen duurzaam moet zijn en dus niet alleen gericht is op het huidige leren, maar ook op het leren in de toekomst. Om die reden moet beoordelen gericht zijn op het testen van hogere-orde vaardigheden die nodig zijn in het beroepenveld en niet enkel binnen de context van de opleiding. Het tweede kenmerk is het ontwikkelen van een geïnformeerd oordeel bij studenten, wat wil zeggen dat niet alleen docenten bepalen wanneer een prestatie aan de criteria voldoet, maar dat ook de student leert om dit zelf te beoordelen. Het derde kenmerk is dat beoordelen moet bijdragen aan het vormen van reflectieve, zelfregulerende studenten in alle fasen van het beoordelingsproces. Daarom moet instructie, leren en beoordelen geïntegreerd worden, zodat de student zich eigenaar voelt van het gehele leerproces. Daarmee lijken het tweede en derde kenmerk uitwerkingen te zijn van hogere orde vaardigheden waar Boud bij kenmerk één naar verwijst. Het vierde kenmerk is dat studenten zichzelf gaan zien als toekomstige beroepsbeoefenaars en dat zij hun prestaties leren beoordelen met behulp van anderen. Dit zorgt ervoor dat misvattingen in het eigen oordeel tijdig gecorrigeerd worden. Om de vertaling van deze kenmerken te concretiseren naar de praktijk beschrijft Boud (2010) acht beoordelingselementen voor de praktijk die voor een beter overzicht worden weergegeven in Tabel 1.

In de Nederlandse context is het concept duurzaam beoordelen opgepakt door Sluijsmans (2008, 2010). Zij identificeert tien kenmerken van duurzaam beoordelen waarin overeenkomsten, maar ook verschillen met het theoretische raamwerk van Boud (2010) te herkennen zijn (zie Tabel 1). Sluijsmans (2010) benoemt drie pijlers voor duurzaam beoordelen: 1) een goed ontwerp van onderwijs waarbij leren en beoordelen samensmelten, 2) heldere communicatie tussen studenten en docenten over het waarom, wat en hoe van beoordelen en 3) het stimuleren van een zelfsturende houding. Anders dan in het model van Boud, benadrukt Sluijsmans dat duurzaam beoordelen afgestemd moet zijn op de 'zone van naaste ontwikkeling' van de unieke student en pleit daarom voor een ipsatief beoordelingsmodel ('ipse' is Latijn voor 'zelf'), waarbij de prestatie van de student wordt vergeleken met eigen prestaties in de tijd en niet met prestaties van andere studenten. Heldere beoordelingscriteria en standaarden die vooraf met de student gecommuniceerd worden en informatierijke feedback, zijn belangrijke voorwaarden voor een ipsatief beoordelingsmodel.

Het model van Fastré, Van der Klink, Sluijsmans & Van Merriënboer (2012) kan worden beschouwd als de volgende stap in de ontwikkeling van duurzaam beoordelen. Dit theoretische model biedt een kapstok om de tot dusver veelal conceptuele uitgangspunten te vertalen naar de onderwijspraktijk. Het model bestaat uit drie noodzakelijke onderdelen voor het realiseren van een

duurzame beoordelingspraktijk, namelijk: voorwaardelijke condities, elementen van duurzaam beoordelen en methoden voor de ontwikkeling van duurzaam beoordelen (zie Figuur 1).



Figuur 1. Geïntegreerd model van Fastré et al. (2012).

Er zijn drie condities voor een duurzame beoordelingspraktijk in dit model. Ten eerste moet er constructive alignment zijn tussen lange-termijn leerdoelen en de beoordeling van deze doelen, ten tweede moeten studenten zichzelf leren beoordelen en leren omgaan met verschillen tussen interne feedback en (externe) feedback van anderen en ten derde moeten studenten de beoordelingscriteria begrijpen en ook zelf kunnen formuleren. Deze condities zijn voorwaardelijk voor een student om zijn/haar duurzame beoordelingsvaardigheid te ontwikkelen. In het model wordt deze complexe vaardigheid verder uitgesplitst in elementen (subvaardigheden) zoals het kunnen beoordelen van de eigen prestatie en het kunnen formuleren van leerdoelen. Tot slot biedt het model methoden voor docenten om duurzame beoordelingsvaardigheden te stimuleren bij studenten, namelijk: modeling examples, coaching (waarmee de auteurs voornamelijk refereren naar het geven van feedback) en scaffolding (de mate van sturing en ondersteuning vanuit de docent moet geleidelijk afnemen). De componenten van dit model zijn in Tabel 1 opgenomen om een vergelijking met de kenmerken van Boud en Sluijsmans mogelijk te maken.

Samenvattend kan worden geconcludeerd dat het werk van Boud (2010), Sluijsmans (2010) en Fastré et al. (2012) hetzelfde doel met betrekking tot duurzaam beoordelen te beschrijven, namelijk; beoordelen moet gericht zijn op het toekomstige leren van de student. De uitgangspunten van duurzaam beoordelen uit deze drie wetenschappelijke bronnen zijn voor een beter overzicht samengevat in Tabel 1, waardoor verschillen en overeenkomsten kunnen worden herkend. In de laatste kolom zijn overlappende uitgangspunten gesynthetiseerd tot generieke ontwerpprincipes, waarmee het theoretische perspectief op duurzaam beoordelen wordt vertegenwoordigd.

Tabel 1

Kenmerken van duurzaam beoordelen volgens Boud (2010), Sluijsmans (2010) en Fastré et al (2012) en de daaruit afgeleide ontwerpprincipes.

Boud	Sluijsmans	Fastré et al.	Ontwerpprincipe
Overeenkomsten			
1) Student heeft een actieve rol in het gehele beoordelingsproces.	1) De lerende eigenaarschap geven over de beoordeling.	Methode 3 – Scaffolding; verantwoordelijkheid voor beoordeling moet geleidelijk naar student verschuiven.	Duurzaam beoordelen vraagt om actieve rol en (toenemende mate van) eigenaarschap van de student in het gehele beoordelingsproces door middel van scaffolding.
2) Authentieke beoordelingssituaties.	2) Betekenisvolle beoordelingstaken waarin leren en beoordelen samensmelten.	Conditie 1 - Constructive alignment tussen lange-termijn leerdoelen en de beoordeling hiervan.	Duurzaam beoordelen vraagt om authentieke en betekenisvolle beoordelingssituaties met constructive alignment tussen lange-termijn leerdoelen en de beoordeling hiervan, zodat instructie, leren en beoordelen geïntegreerd zijn.
4) Actief keuzes maken in eigen leerproces door middel van een geïnformeerd oordeel.	3) Het stimuleren van zelfsturing.	Elementen - Stimuleren van alle elementen van de complexe duurzame beoordelingsvaardigheid van de student.	Duurzaam beoordelen is gericht op stimulering van alle elementen van de complexe zelfsturende beoordelingsvaardigheid van de student waardoor de student een geïnformeerd oordeel ontwikkelt.
5) Gebruik van models en worked examples als voorbeelden van good practice. Student moet na beoordeling de kans krijgen om werk te verbeteren om groei te ervaren.	4) Beoordelen is start van een nieuwe leercyclus.	Methode 2 – Docenten maken gebruik van modeling examples.	Duurzaam beoordelen is de start van een nieuwe leercyclus waarin de student de kans krijgt zijn voorgaande prestatie te verbeteren. Bij duurzaam beoordelen worden modeling examples gebruikt om de student te laten zien wat van ze verwacht wordt.

-	5) Heldere standaarden en criteria formuleren.	Conditie 3 - Student criteria begrijpen en zelf kunnen formuleren.	Duurzaam beoordelen vraagt om heldere beoordelingscriteria en moet de student zelf beoordelingscriteria kunnen formuleren.
6) Student moet leren feedback geven en ontvangen.	6) Beoordeling moet informatierijke feedback opleveren.	Methode 1 – Coaching door de docent door middel van informatierijke feedback.	Bij duurzaam beoordelen geeft en ontvangt de student en de docent informatierijke feedback.
7) Student moet eigen prestatie leren beoordelen met behulp van anderen (peer-assessment).	-	Conditie 2: Student maakt bewuste vergelijking tussen interne en externe feedback.	Bij duurzaam beoordelen leert de student zijn eigen prestatie beoordelen op basis van zelfreflectie en externe feedback.
Verschillen			
8) Student betrekken in het ontwerpen van beoordelingstaken.	-	-	Bij duurzaam beoordelen wordt de student betrokken in het ontwerpen van beoordelingstaken.
-	7) Mate van duurzaamheid van kennis erkennen.	-	Bij duurzaam beoordelen wordt de mate van duurzaamheid van kennis erkend.
-	8) Formeel en non-formeel leren erkennen.	-	Bij duurzaam beoordelen wordt formeel en non-formeel leren erkend.
-	9) De unieke student staat centraal door middel van een ipsatief beoordelingsmodel.	-	Bij duurzaam beoordelen staat de unieke student centraal door middel van een ipsatief beoordelingsmodel.
-	10) Accent leggen op leren voor de toekomst.	-	<i>Geen ontwerpprincipes maar het doel van duurzaam beoordelen.</i>

Noot: Acht beoordelingselementen van Boud (2010), tien kenmerken van duurzaam beoordelen van Sluijsmans (2010) en de zeven componenten van het model van Fastré et al. (2012) en de daaruit gesynthetiseerde ontwerpprincipes.

In de reviewstudie van Boud en Soler (2016) naar onderzoeken over duurzaam beoordelen sinds de introductie van het concept, concluderen zij dat de duurzame functie van beoordelen een zinvolle aanvulling is in de wetenschappelijke discussie omtrent beoordelen en een interessant uitgangspunt biedt voor veranderingen in de beoordelingspraktijk. Anderzijds erkennen zij ook dat er nog veel onderzoek nodig is om de vertaalslag van theoretische modellen zoals die van Fastré et al. naar onderwijspraktijk te realiseren. Fastré en collega's (2012) noemen zelf ook een belangrijke voorwaarde voor het realiseren van een duurzame beoordelingspraktijk, namelijk: dat docenten en beleidsmakers het belang van het ontwikkelen van duurzame beoordelingsvaardigheden voor studenten moeten inzien (Fastré et al., 2012). Het is daarom van belang om opvattingen van docenten met betrekking tot duurzaam beoordelen te inventariseren voordat verdere implementatie kan worden gerealiseerd.

1.1.2 Opvattingen van docenten over beoordelen

Opvattingen van docenten over beoordelen vormen een belangrijke basis voor verdere implementatie van de in de literatuur geformuleerde uitgangspunten omtrent duurzaam beoordelen (Fastré et al., 2012; Sluijsmans, 2010). Om onderwijsvernieuwingen te laten slagen is het van belang dat naast het ontwikkelen van specifieke kennis en uitvoerende vaardigheden, ook rekening wordt gehouden met de opvattingen van docenten (Thurlings & Van Diggelen, 2019; Trigwell, Prosser, & Taylor, 1994; Verloop et al., 2001). Een groot deel van de onderwijsinnovaties is niet succesvol, omdat docenten blijven vasthouden aan of terugkeren naar oude routines (Verloop et al., 2001). Het besef van de invloed van opvattingen van docenten op de onderwijspraktijk en onderwijsvernieuwingen is vrij recent en diverse onderzoeken tonen aan dat om veranderingen in de onderwijspraktijk te bewerkstelligen de opvattingen van docenten als uitgangspunt genomen moet worden (Birman, Desimone, Porter, & Garet, 2000; Calderhead, 1996; Hawley & Valli, 1999). Dit geldt ook voor veranderingen in de beoordelingspraktijk van docenten. Onderzoek toont aan dat de beoordelingspraktijk van docenten, beïnvloed wordt door hun opvattingen over beoordelen (Calderhead, 1996; Pajares, 1992; Sadler, 1989; Thomas, 2012; Verloop et al., 2001). Dit betekent niet dat opvattingen van docenten de standaard van onderwijsontwikkelingen moeten zijn, maar ze moeten wel het startpunt vormen om succesvolle interventies en innovaties te bewerkstelligen (Brand-Gruwel et al., 2019; Verloop et al., 2001).

In de literatuur omtrent opvattingen van docenten worden verschillende termen en definities gehanteerd, waarbij sommigen termen onderscheid maken tussen kennis en overtuigingen (beliefs). In navolging van Thompson (1992) en Brown (2004) worden opvattingen (conceptions) in dit onderzoek gedefinieerd als “a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like” (Thompson, 1992, p. 130). In deze definiëring worden zowel kennis als overtuigingen in één construct gecombineerd.

Een beperkt aantal onderzoeken zijn uitgevoerd naar de opvattingen van docenten met betrekking tot formatief beoordelen. In het onderzoek van Thomas, Deaudelin, Desjardins en Dezutter (2011) onder 13 Canadese basisscholen, zijn docentenopvattingen over de tijd, vorm en rol van de docent/student in kaart gebracht. Deze docenten beschreven formatief beoordelen als integraal in het leerproces en identificeerden verschillende aanpakken om kennisgebreken in het leerproces van studenten te achterhalen. Over de rol van de docent liepen de opvattingen van docenten uiteen, maar de meerderheid zag het formatieve beoordelingsproces als een verantwoordelijkheid van de docent. Het gebruik van self- en peerassessmentstrategieën bleek ook minimaal ingezet te worden in de klas (Thomas et al., 2011). Het onderzoek van Davis en Neitzel (2011) bevestigt dat docenten verantwoordelijkheid voor het beoordelingsproces veelal bij zichzelf leggen. De meeste docenten zien het primaire doel van formatieve beoordelingen als een manier om de competenties en motivatie van studenten te identificeren, maar rapporteerden ook moeilijkheden met het implementeren van formatieve beoordelingsstrategieën in de klas (Davis & Neitzel, 2011). Er zijn geen onderzoeken bekend naar opvattingen van docenten over duurzaam beoordelen.

Als opvattingen van docenten het startpunt vormen voor onderwijsinnovatie, is de belangrijkste vraag niet hoe inzichten vanuit wetenschappelijke evidentie het beste overgebracht kunnen worden op docenten, maar hoe er een dialoog kan ontstaan tussen wetenschappelijke inzichten en docenten inzichten. Er moet een uitwisseling ontstaan tussen de theoretische ontwerpprincipes uit de wetenschappelijke literatuur en docenten ervaringen en opvattingen, waarbij beide vormen van input elkaar complementeren (Verloop et al., 2001). Inzicht in de opvattingen van de mbo-docenten van de participerende onderwijsinstelling is van belang om ontwerpprincipes te kunnen formuleren die de goedkeuring van de betrokkenen hebben (Brand-Gruwel et al., 2019). Het huidige ontwerpgerichte onderzoek heeft daarom als doel om ontwerpprincipes te formuleren waarbij de in de vorige paragraaf geformuleerde theoretische ontwerpprincipes worden gecomplementeerd met opvattingen van docenten.

1.2 Vraagstellingen en hypothesen

Duurzaam beoordelen biedt een nieuw perspectief om de bestaande beoordelingspraktijk in het mbo te hervormen naar een praktijk die beter aansluit bij de doelstellingen van een leven lang leren. Op dit moment ontbreekt het echter nog aan concrete en docentgedragen ontwerpprincipes om deze onderwijsvernieuwing succesvol te implementeren. Doel van dit onderzoek is daarom om ontwerpprincipes te formuleren waarbij theoretische ontwerpprincipes en docenten opvattingen elkaar complementeren. De centrale onderzoeksvraag is:

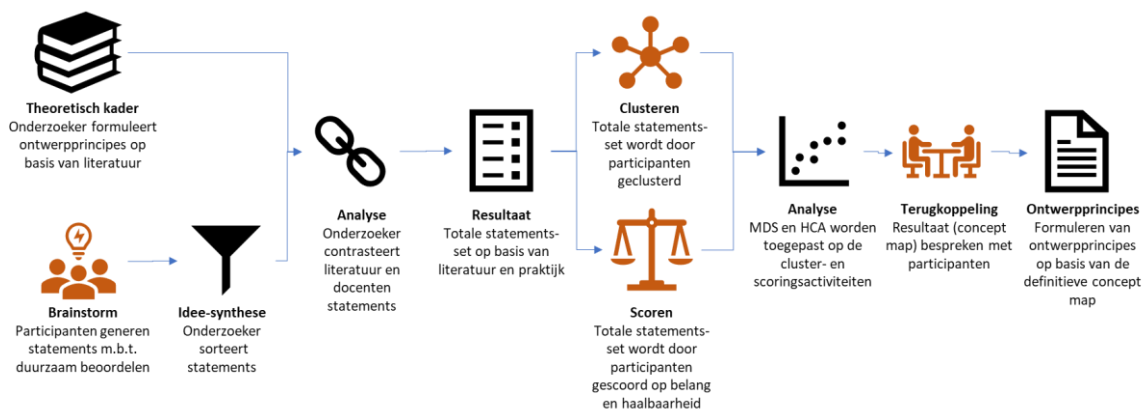
Welke ontwerpprincipes ter bevordering van duurzaam beoordelen kunnen worden geformuleerd op basis van opvattingen van docenten en de wetenschappelijke literatuur? De bevindingen uit dit ontwerpgerichte exploratieve onderzoek komen voort uit de specifieke context van de participerende mbo-instelling, waardoor de resultaten moeilijk generaliseerbaar zullen zijn.

2. Methode

2.1 Ontwerp

Om tot ontwerpprincipes te komen op basis van opvattingen van docenten en de wetenschappelijke literatuur is een mixed-methods design gebruikt, waarbij zowel kwantitatieve als kwalitatieve methoden van dataverzameling gecombineerd werden. In dit onderzoek werd de methode group concept mapping (GCM) gebruikt om opvattingen van docenten omtrent duurzaam beoordelen te inventariseren. GCM komt voort uit de traditie van community-based participatory onderzoek en omarmt de kennis, ervaringen, behoeftes en meningen van degene die het meest met het onderwerp van doen hebben (group wisdom). Deze methode is geschikt om de opvattingen van een groep te verzamelen om tot een conceptueel framework te komen, waaruit acties kunnen worden afgeleid (Rosas & Kane, 2018). Een voordeel van dit ontwerp is dat de deelnemers in verschillende fasen van het onderzoek werden betrokken. Daarmee biedt dit ontwerp een bottom-up benadering wat bijdraagt aan draagvlak voor de implementatie-bevorderende ontwerpprincipes die uit het onderzoek zijn afgeleid. Door ontwerpprincipes uit de literatuur toe te voegen aan de docentgegenereerde statements-set biedt GCM ook de mogelijkheid om stellingen uit de praktijk en theorie te integreren in het onderzoek. Een nadeel van deze methode is dat er gedurende het proces veel keuzes gemaakt werden door de onderzoeker over het meest geschikte kwalitatieve data (Creswell, 2014). Desondanks is GCM een goed gedocumenteerde en een gevestigde onderzoeksmethodologie waarin de voordelen van zowel kwalitatieve als kwantitatieve data-analyse gecombineerd wordt, die reeds is toegepast in honderden gepubliceerde artikelen (Rosas & Kane, 2018).

Voor de beschrijving van onderzoeken uitgevoerd met GCM zijn (nog) geen eenduidige standaarden beschikbaar (Rosas & Kane, 2012; Trochim & McLinden, 2017). De rapportage van GCM wordt in recente studies gedaan aan de hand van de stappen van GCM. De stappen van het huidige onderzoek zijn gevisualiseerd in Figuur 2, waarin zwarte acties uitgevoerd zijn door de onderzoeker (zie data-analyse) en rode acties uitgevoerd zijn door participanten (zie procedure).



Figuur 2. Schematische weergave onderzoeksontwerp waarin zwarte acties uitgevoerd worden door de onderzoeker en rode acties uitgevoerd worden door participanten.

Allereerst leverde het theoretische kader ontwerpprincipes op met betrekking tot duurzaam beoordelen op basis van de wetenschappelijke literatuur. Om opvattingen vanuit het praktijkperspectief van docenten te achterhalen vond er een online brainstorm plaats, waarbij participanten gevraagd werden stellingen met betrekking tot duurzaam beoordelen te genereren door te reageren op een focus prompt. Deze stellingen werden door de onderzoeker gesorteerd tijdens de idee-synthese fase. Daarna werden deze docenten-statements gecontrasteerd met de theoretische ontwerpprincipes. Deze analyse leverde een statements-set op waarin stellingen uit de praktijk werden aangevuld met de ontwerpprincipes uit de literatuur. Vervolgens sorteerden en scoorden de participanten alle stellingen uit de statements-set. Op basis van deze sorteer- en scoringsactiviteiten werden cluster-analyses uitgevoerd. Tot slot werden de onderzoeksresultaten teruggekoppeld aan de deelnemers om de cluster- en label interpretaties van de onderzoeker te verifiëren (Rosas & Kane, 2018). Op basis de resulterende clusters in de conceptmap werden nieuwe ontwerpprincipes geformuleerd waarin zowel opvattingen van docenten als ontwerpprincipes uit de literatuur zijn meegenomen.

2.2 Participanten

Het onderzoek werd uitgevoerd op een mbo-instelling in het Zuiden van Nederland, vanwege de ambitie van deze onderwijsinstelling om een duurzame beoordelingspraktijk te realiseren. Het betrof een middelgrote vakschool die creatief-technische opleidingen aanbiedt op niveau 3 en 4 met 216 medewerkers. Enkel onderwijzend personeel ($n = 170$) werd voor het onderzoek benaderd. De percentuele verdeling tussen mannelijke en vrouwelijke docenten binnen deze instelling was 45/55 en de leeftijd varieerde tussen de 20 en 65 jaar met een gemiddelde leeftijd van 46 jaar. Het was belangrijk voor het draagvlak van de ontwerpprincipes dat deelnemers een goede afspiegeling vormden van de verschillende opleidingen en leerjaren binnen de organisatie, zodat mogelijke cultuurverschillen binnen de teams konden worden meegenomen en dat alle teams zich gehoord en betrokken voelden. Het was daarom van belang dat vanuit ieder onderwijsteam ($n = 15$) minimaal één docent deelnam aan de brainstormfase van het onderzoek. Het was voor het onderzoek niet noodzakelijk dat dezelfde participanten aan alle fasen van het onderzoek deelnemen, hoewel de uitkomsten van de GCM beter begrepen worden als dit wel het geval is (Trochim, 1989).

Kane hanteert in haar webinar (2020) de volgende richtlijn voor participanten aantallen in de verschillende onderzoeksfasen. Voor de brainstormfase zijn de aantallen volledig afhankelijk van de omvang van het onderzoek en een groep van 10 tot 20 deelnemers biedt al een variëteit aan meningen. In het huidige onderzoek hebben 27 docenten input geleverd aan de brainstormfase (16%). Voor de sorteer en scoringsfase van het onderzoek adviseren Rosas en Kane een minimum van 10 deelnemers, waarbij voor de sorteerfase een optimum waarin verzadiging optreedt ligt tussen de 25 en 40 deelnemers. In dit onderzoek hebben 20 participanten een bijdrage geleverd aan de sorteerfase en 11 participanten aan de scoringsfase. Om de resultaten van de sorteerfase beter in context te kunnen plaatsen hebben de deelnemers aan deze fase demografische gegevens ingevuld. Hieruit bleek dat van

de 15 onderwijsteams, 10 teams vertegenwoordigd werden in deze fase van het onderzoek door één of meerdere docenten. De gemiddelde leservaring van deze docenten lag op 14,9 jaar. Zes van de twintig docenten verzorgden theorielessen en de overige 14 docenten verzorgden praktijklessen.

2.3 Materialen

Voor de uitvoering van de dataverzameling is gebruik gemaakt van de software Concept System Global MAX™ van Concept Systems, Inc. Dit online programma omvat verschillende software tools waarmee het gehele GCM proces kan worden doorlopen, zoals: het aanbieden van een focus prompt, het toevoegen en beheersen van participanten, het creëren van een database voor de door deelnemers gegenereerde stellingen, het sorteren en scoren van de stellingen door participanten, het uitvoeren van de concept mapping analyse en het genereren van de resultaten (Concept System Incorporated, 2017).

Voor de brainstormfase is een focusprompt gebruikt om docenten stellingen te laten generen. Ter inleiding van de focusprompt kregen de deelnemers de volgende instructie:

“Duurzaam beoordelen is een concept dat is bedacht om aan te geven dat de beoordelingspraktijk in het onderwijs studenten zou moeten voorbereiden op een leven lang leren. Dit betekent dat beoordelingen niet alleen gericht moeten zijn op het beoordelen van opleidings- en beroepsgerichte vaardigheden, maar ook vaardigheden die de student nodig heeft om zelfstandig en zijn hele leven lang (dus ook na de opleiding) richting te kunnen geven aan zijn eigen persoonlijke en professionele ontwikkeling. Om naar een meer duurzame manier van beoordelen te gaan in het dagelijkse onderwijs is het van belang om jouw opvattingen over belangrijke en haalbare uitgangspunten omtrent duurzaam beoordelen te achterhalen. Daartoe willen we je vragen om jouw opvattingen ten aanzien van een duurzame manier van beoordelen te delen in deze online brainstorm.” De respondenten vullen daarbij de volgende focusprompt aan: “Voor het realiseren van een duurzame beoordelingspraktijk in het dagelijks onderwijs waarmee we de student helpen voorbereiden op een leven lang leren is het van belang om.....”

2.4 Procedure

De procedure beschrijft hoe het onderzoek is uitgevoerd vanaf eerste contact tot het laatste contact met de deelnemers van het onderzoek.

Werving: Voor de werving van deelnemers is met behulp van de communicatieafdeling een oproep tot deelname uitgezet op de interne communicatiepagina (Sharepoint) van de onderwijsinstelling en de online nieuwsbrief (Outlook). Geïnteresseerden werden verzocht de informatiebrief te lezen en vervolgens de online toestemmingsverklaring te ondertekenen en te mailen naar de onderzoeker, conform de procedures van de Ethische richtlijnen. Ook zijn de opleidingscoördinatoren van de 15 onderwijsteams via de locatiemanagers per mail benaderd met het verzoek om de oproep tot deelname binnen hun onderwijsteam onder de aandacht te brengen. In totaal hebben

30 deelnemers een toestemmingsverklaring voor deelname aan (één of meerdere fases) van het onderzoek ondertekend en naar de onderzoeker gemaild.

Brainstormfase: Deze 30 deelnemers hebben per mail een uitnodiging ontvangen om deel te nemen aan de brainstormfase via een link naar de webgebaseerde omgeving van Content System Global. Als respondenten op de website kwamen kregen zij een korte introductie en de focus prompt voorgelegd met de vraag om deze in eigen woorden aan te vullen. Respondenten konden meerdere stellingen aandragen. In totaal zijn er 87 stellingen gegenereerd.

Sorteren: In deze fase zijn 28 deelnemers op basis van de toestemmingsverklaring uitgenodigd via een link in de mail om de stellingen van de totale statements-set ($n = 84$) te clusteren in de webgebaseerde omgeving van Content Systems Global. De respondenten kregen de volgende instructie in de online omgeving: “Cluster de stellingen die volgens jou bij elkaar horen op basis van de betekenis of per onderwerp. Geef ieder cluster een naam die het onderwerp of de inhoud het beste beschrijft. Je kunt de stellingen sorteren door ze te slepen. Je begint met de eerste stelling en je sleept het naar het midden van de pagina. Het systeem vraagt je een naam te geven voor het cluster waarin je de stelling wilt plaatsen. Het cluster wordt aangemaakt en de stelling staat erin. Je sleept de tweede stelling naar het midden en maakt zo een nieuw cluster of je zet het in een bestaand cluster.” De richtlijnen die alle deelnemers meekrijgen zijn: maak geen clusters met als naam “belangrijk”, “moeilijk”, “anders” of “varia” en plaats geen twee stellingen in twee clusters tegelijkertijd. Elke stelling komt voor in één cluster.

Scoring: In de scoringsfase zijn op basis van de toestemmingsverklaring 21 respondenten per mail uitgenodigd om de statements-set te scoren via de online tool. Alle stellingen uit de set zijn in random volgorde geplaatst in de webgebaseerde omgeving van Content Systems Global. De respondenten werden gevraagd antwoord te geven op de vraag “Op een schaal van 1 tot 5, hoe *belangrijk* vind je de volgende stellingen omtrent duurzaam beoordelen?” en de vraag “Op een schaal van 1 tot 5, hoe *haalbaar* vind je de volgende stellingen omtrent duurzaam beoordelen?” Aan alle respondenten werd dus gevraagd iedere stelling twee keer te scoren (op belang en op haalbaarheid) op een vijfpuntsschaal, lopend van 1 (helemaal niet belangrijk) tot en met 5 (heel belangrijk) en van 1 (geheel niet haalbaar) tot 5 (heel erg haalbaar). Het gebruik van een vijfpuntsschaal is een bewuste keuze om de participanten de optie te geven om neutraal te zijn ten opzichte van een stelling. Een even schaal dwingt een stellingname af, maar daardoor kan juist ook weerstand optreden (Creswell, 2014).

Terugkoppelingssessie: Nadat de data-analyse resulteerde in een concept map, zijn de resultaten teruggekoppeld aan de participanten in een online bijeenkomst waarvoor zij per mail werden uitgenodigd door de onderzoeker. Tijdens deze bijeenkomst werd een korte inhoudelijke toelichting gegeven op het onderzoek en het onderzoeksproces, zodat participanten zicht kregen op het aandeel dat zij geleverd hebben aan het onderzoek en de huidige status van het onderzoek. Aan de 8 aanwezige participanten werd gevraagd of de resultaten helder zijn, of zij de resultaten verrassend of voor de

hand liggend vinden en of de resultaten en de gekozen clusterlabels passend vinden bij de clusteroplossing (Rosas & Kane, 2018).

2.5 Data-analyse

GCM is een iteratief proces, waarbij momenten van dataverzameling werden afgewisseld met momenten van data-analyse. Deze paragraaf beschrijft de in het zwart gevisualiseerde stappen van het onderzoeksproces van Figuur 2.

Idee-synthese: In deze fase zijn de tijdens de brainstorm gegenereerde stellingen door de onderzoeker samen met een tweede beoordelaar gecontroleerd op begrijpelijkheid, eenduidigheid en overlap. Inhoudelijk overlappende stellingen zijn verwijderd. Het was van belang dat iedere unieke stelling vertegenwoordigd werd in de statements-set, maar niet hoe vaak iets is gezegd.

Totale statements-set: Nadat de idee-synthese fase een statements-set had opgeleverd met enkel docentenstellingen, zijn de 12 ontwerpprincipes uit de literatuur door de onderzoeker toegevoegd aan de statements-set. Stellingen uit de literatuur werden gelabeld met L1, L2 etc. en stellingen uit de brainstorm werden gelabeld met B1, B2, etc. Al deze stellingen zijn onder elkaar in een tabel geplaatst en vervolgens werden inhoudelijk overlappende stellingen tussen literatuur en praktijk samengevoegd waarbij de ID-codes van de oorspronkelijke stellingen zichtbaar blijven. Als richtlijn wordt gesteld dat een statements-set van 100 of minder het best hanteerbaar is (Rosas & Kane, 2012). De stellingen zijn vervolgens in een random volgorde teruggezet in de tool, zodat de participanten deze stellingen konden sorteren in clusters en scoren op belang en haalbaarheid.

Multidimensional scaling: Het volgende moment van data-analyse vond plaats nadat de stellingen door de participanten waren geclusterd en gescoord. Participanten moesten minstens 75% van alle stellingen ($n = 63$) hebben gesorteerd om te worden goedgekeurd (Concept Systems Incorporated, 2017). Voorafgaand aan MDS werden de sorteeractiviteiten van de deelnemers weergegeven als individuele matrices en vervolgens gecombineerd in een similarity matrix (Rosas & Kane, 2018). De individuele matrix is binair, want er komt een '1' in een cel als twee stellingen samen zijn geclusterd en een '0' als dat niet het geval is. Het aantal rijen en kolommen van de matrix komt overeen met het aantal gegenereerde stellingen en is daarom symmetrisch. In de similarity matrix bevat elke cel het aantal personen dat de combinatie van twee stellingen in dezelfde groep heeft geplaatst tijdens het sorteren en is dus een opsomming van de individuele binaire matrices. MDS brengt de similarity matrix als een tweedimensionale weergave in kaart door twee coördinaten te berekenen die de afstand representeert van een stelling tot de andere stellingen. Door dit voor alle stellingen te doen ontstaat er een puntenwolk. De mate waarin de puntenwolk de similarity matrix representeert wordt uitgedrukt als een stress-waarde en deze zou tussen de .20 en de .36 moeten vallen (Kruskal & Wish, 1978; Rosas & Kane, 2018).

Hiërarchische clusteranalyse: Om tot clusters te komen op basis van de puntenwolk is vervolgens een hiërarchische clusteranalyse (HCA) uitgevoerd die gebaseerd is op Ward's algoritme. Hierbij zijn

de waarden van de coördinaten uit de twee dimensionale MDS opgedeeld in vlakken die aanpalend zijn, maar niet met elkaar overlappen (Kane & Trochim, 2007). Om de relatieve samenhang van stellingen en het onderscheidend vermogen van de clusters gebaseerd op de puntenwolk beter te beoordelen, is anchoring en bridging analysis toegepast. Dit leverde per stelling een waarde op tussen 0 en 1, waarbij een waarde dicht bij 0 (een anchor) aangeeft dat er een sterkere relatie is tussen deze stelling en de omliggende stellingen en een waarde dichtbij 1 (een bridge) een verspreide relatie indiceert. Stellingen met een lage brugwaarde geven aan dat deze veelal gelijk gesorteerd zijn door alle docenten en zijn een betere indicator voor dat deel van de puntenwolk. De brugwaarde van het cluster geeft aan hoe consistent en passend de stellingen binnen het cluster zijn en geeft inzicht in hoe de clusters onderling gerelateerd zijn. Om de samenhang van items per cluster in beeld te brengen, is een gemiddelde brugwaarde per cluster weergegeven, aangevuld met spreidingsmaten zoals variantie, minimum, maximum en mediaan. Een spanningsanalyse geeft aan met welke stellingen uit andere clusters een stelling vaak samen is gesorteerd. De dikte van de lijnen geeft de mate aan hoe vaak de stelling met stellingen buiten het cluster is gesorteerd.

De resultaten van de HCA en de anchoring en bridging analysis werden als uitgangspunt gebruikt om tot een passend aantal clusters te komen, in combinatie met de inhoudelijke interpretatie van de stellingen in een cluster. De uitkomst van de HCA liet clusteroplossingen zien van 15 tot 4 clusters. Startend bij de hoogste aantal clusters werd bij iedere reductie gekeken of een samenvoeging van stellingen inhoudelijk zinvol is (Trochim, 1989). Bij de stap van 8 naar 7 clusters werden twee clusters die betrekking hebben op verschillende thema's samengevoegd. Een samenvoeging van deze stellingen was dus inhoudelijk niet zinvol, waardoor in dit onderzoek gekozen is voor een 8-clusteroplossing. Trochim (1989) benoemt dat een clusteroplossing nooit optimaal is. De afstand tussen de stellingen is betekenisvoller dan het cluster waarin deze is geplaatst. De puntenwolk is daarom leidend en niet de gekozen clusterindeling.

Om tot clusterlabels te komen is samen met twee andere beoordelaars overeenstemming gezocht die inhoudelijk de stellingen in ieder cluster het beste representeren. Hierbij zijn de stellingen met de laagste brugwaarde als uitgangspunt genomen, omdat zij het cluster het best representeren. Daarnaast is getracht zoveel mogelijk bij de originele stem van de respondenten aan te sluiten door naar de labels te kijken die de respondenten bij het sorteren hebben aangedragen (Trochim & McLinden, 2017). Ook hebben de participanten tijdens de terugkoppelingssessie reactie geven op de gekozen clusterlabels waarna de definitieve concept map en clusterlabels zijn geformuleerd.

3. Resultaten

3.1 Resultaten sorteerddata

De verschillende momenten van data-analyse uit de vorige paragraaf leverde de volgende resultaten op met betrekking tot de sorteerddata:

Idee synthese. In totaal hebben 27 deelnemers 87 stellingen gegenereerd en deze zijn door de onderzoeker met een tweede beoordelaar teruggebracht in aantal. Daarbij zijn de volgende stappen doorlopen: 1) de stellingen zijn waar nodig taalkundig aangepast om beter bij de focus prompt aan te sluiten en spelfouten zijn gecorrigeerd om de leesbaarheid te vergroten, 2) er zijn 11 vage en onduidelijke stellingen verwijderd, 3) er zijn 8 multifocus stellingen opgesplitst waardoor er 12 additionele stellingen zijn toegevoegd en 4) er zijn 15 dubbele stellingen verwijderd waardoor een totale aantal stellingen uitkwam op $n = 73$.

Totale statements-set. Vervolgens zijn de 12 ontwerpprincipes uit de literatuur door de onderzoeker toegevoegd aan de statements-set. Slechts één ontwerpprincipe uit de literatuur vertoonde voldoende overlap met een stelling uit de brainstorm om als dubbel te worden beschouwd, namelijk: Duurzaam beoordelen vraagt om authentieke en betekenisvolle beoordelingssituaties met constructieve alignment tussen lange-termijn leerdoelen en de beoordeling hiervan, zodat instructie, leren en beoordelen geïntegreerd zijn. Er is voor gekozen om de stelling in de bewoording van de docent te hanteren, en de stelling uit de literatuur te verwijderen uit de totale statements-set. Dit leverde een uiteindelijke statements-set op van $n = 84$ bestaande uit stellingen die enkel in de literatuur voorkomen, stellingen die enkel in de brainstorm genoemd zijn en één stelling die zowel in de literatuur als in de brainstorm genoemd werd (zie Bijlage B).

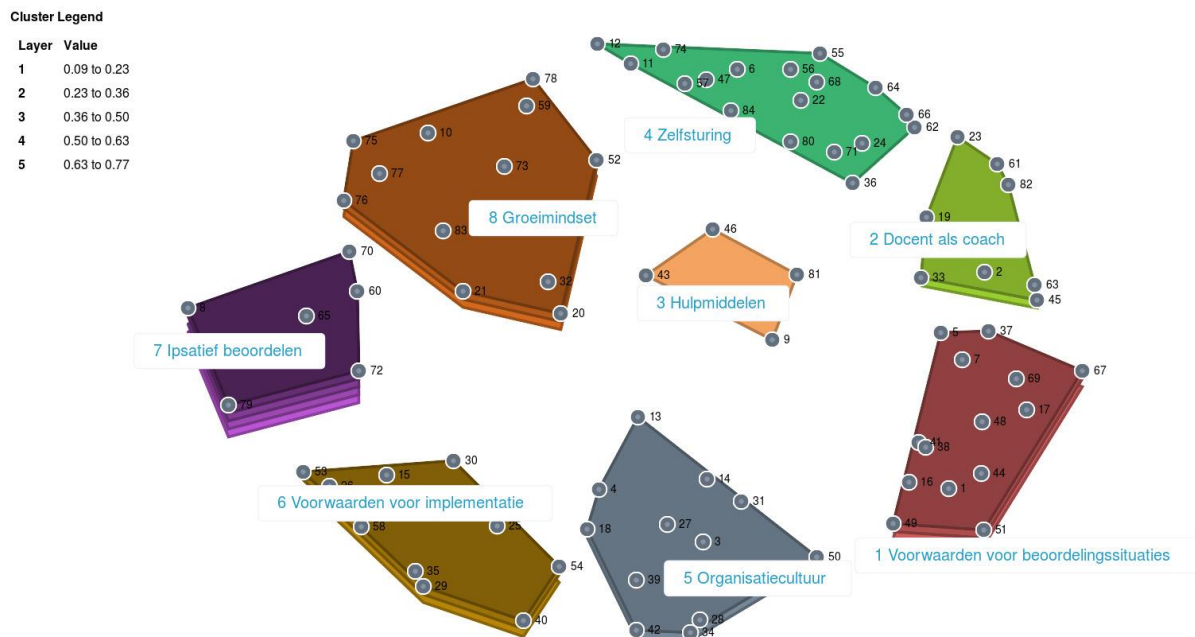
Resultaten MDS-analyse. In totaal hebben 20 participanten de 84 stellingen gesorteerd. Hiervan zijn van 19 participanten de sorteeractiviteiten goedgekeurd. De MDS-analyse van de data resulteerde in een twee-dimensionele puntenwolk (Figuur 3).



Figuur 3. Puntenwolk gegenereerd vanuit de MDS-analyse op basis van 19 participanten, met 10 iteraties en een stress-index van 0.3210.

De stress-index voor de puntenwolk zoals gevisualiseerd in Figuur 3 is 0.321 en kan dus worden beschouwd als een redelijk goede fit. Dit betekent dat de puntenwolk de similarity-matrix goed representeert. De kleuren in Figuur 3 geven de brugwaarde van iedere stelling weer. Datapunten zonder kleur hebben een brugwaarde tussen de 0 en 0.20, paars tussen de .20 en .40, groen tussen de .40 en .60, geel tussen de .60 en .80 en rood tussen .80 en 1.00. De stellingen rechtsboven de puntenwolk zijn dus veelal gelijk gesorteerd door de docenten, want zij hebben een lage brugwaarde. Stellingen links in de puntenwolk hebben juist een hoge brugwaarde en zijn dus niet vaak samen gesorteerd.

Resultaten HCA. De Hiërarchische Cluster Analysis (HCA) geeft aan hoeveel clusters er kunnen worden gevormd op basis van de puntenwolk, uitgaande van de hoeveelheid clusters die de 19 respondenten hebben gecreëerd tijdens het sorteren van de 84 stellingen. De participanten hebben de stellingen gemiddeld in 7,6 clusters gesorteerd ($SD = 3.48$; range: 4-15; Mdn: 7). De analyse liet zien dat een oplossing met acht clusters het beste bij de data past. Bij de stap van 8 naar 7 clusters werden twee clusters inhoudelijk verschillende thema's samengevoegd, waardoor gekozen is voor een 8-clusteroplossing die in Figuur 4 wordt weergegeven.



Figuur 4. De 8-cluster oplossing met labels.

Stellingen met een lage brugwaarde representeren het cluster het best en een lage gemiddelde brugwaarde geeft aan dat het cluster bestaat uit stellingen die op dezelfde wijze door participanten zijn gesorteerd. In Figuur 4 is de gemiddelde brugwaarde van ieder cluster doormiddel van lagen gevisualiseerd. De labels voor ieder cluster zijn samen met twee andere beoordelaars bepaald door te kijken naar de items met de laagste brugwaarde en door inhoudelijk naar de stellingen in het cluster te kijken. Tijdens de terugkoppelingssessie hebben 8 participanten gereageerd op de gekozen clusteroplossing en clusterlabels wat heeft geleid tot de volgende clusterlabels: 1) *Voorwaarden voor beoordelingssituaties*, 2) *Docent als coach*, 3) *Hulpmiddelen*, 4) *Zelfsturing*, 5) *Organisatiecultuur*, 6) *Voorwaarden voor implementatie*, 7) *Ipsatief beoordelen*, 8) *Groeimindset*.

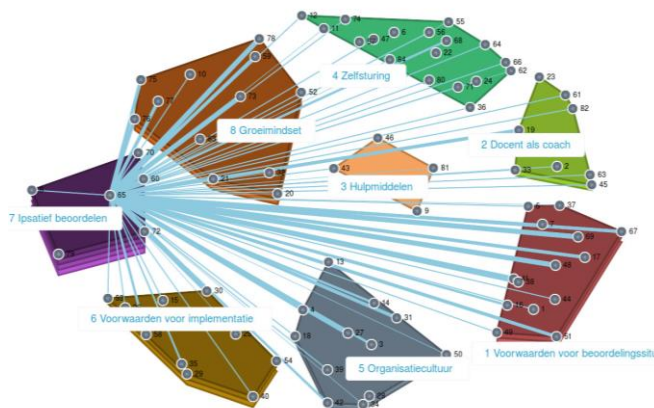
Tabel 2 geeft een overzicht van de acht clusters met bijbehorende statistieken, zoals gemiddelde brugwaarde, standaard deviatie, variantie, minimum, maximum en mediaan.

Tabel 2

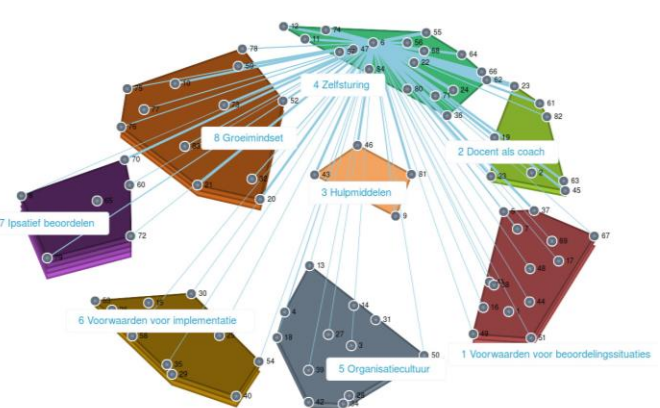
Acht clusters met aantal stellingen per cluster, gemiddelde brugwaarde, variantie, minimum, maximum en mediaan.

Cluster	Aantal stellingen	Gemiddelde brugwaarde (SD)	Variantie	Minimum	Maximum	Mediaan
1 Beoordelingsituaties	14	0.38 (.06)	0.004	0.28	0.53	0.36
2 Docent als coach	8	0.32 (.08)	0.006	0.15	0.43	0.33
3 Hulpmiddelen	4	0.17 (.02)	0.000	0.16	0.20	0.17
4 Zelfsturing	18	0.09 (.07)	0.005	0.00	0.30	0.09
5 Organisatiecultuur	12	0.25 (.03)	0.001	0.22	0.33	0.25
6 Implementatie	10	0.40 (.10)	0.009	0.28	0.57	0.36
7 Ipsatief beoordelen	6	0.77 (.13)	0.017	0.57	1.00	0.76
8 Groeimindset	12	0.45 (.15)	0.023	0.25	0.73	0.42

Uit Tabel 2 valt af te lezen dat het cluster *Zelfsturing* de laagste gemiddelde brugwaarde heeft en dus het meest consistent is. De stellingen binnen dit cluster zijn vaak samen gesorteerd en hebben minder verwantschap met de stellingen uit de andere clusters. Het cluster *Ipsatief beoordelen* heeft de hoogste gemiddelde brugwaarde en de stellingen in dit cluster zijn dan ook vaak gesorteerd met stellingen uit andere clusters. De spanningsanalyse in Figuur 5 en 6 maakt dit inzichtelijk.



Figuur 5. Spanningsanalyse van een stelling uit het cluster Ipsatief beoordelen met een hoge brugwaarde.



Figuur 6. Spanningsanalyse van een stelling uit het cluster Zelfsturing met een lage brugwaarde.

In Figuur 5 is te zien dat de stelling ‘te focussen op groei’ uit het cluster *Ipsatief beoordelen* een brugwaarde van 0.8 heeft en dikke lijnen naar stellingen uit het cluster *Voorwaarden voor*

beoordelingssituaties heeft lopen. In Figuur 6 is te zien dat de stelling ‘de student leert zijn eigen prestatie beoordelen op basis van zelfreflectie en externe feedback’ met een lage brugwaarde van 0.016 dunne lijnen heeft lopen naar stellingen uit andere clusters. In Bijlage C is het volledige clusterrapport te vinden met een uitgebreid overzicht van alle items in ieder cluster, de brugwaarde per item en de statistieken per cluster. In het clusterrapport is ook te zien dat de theoretische ontwerpprincipes gelijkmatig over de acht clusters verdeeld zijn (zie dikgedrukte stellingen in het clusterrapport van bijlage C). De individuele clusters met de items met laagste brugwaarde worden nu kort besproken.

Items per cluster. Het cluster *Voorwaarden voor beoordelingssituaties* (n = 14) heeft een gemiddelde brugwaarde van 0.38 en is dus redelijk consistent. Het draait in dit cluster om kenmerken en voorwaarden voor duurzame beoordelingssituaties. Het bevat stellingen als ‘de student zich veilig laten voelen’, ‘het voor de student praktisch maken’ en ‘de student betrekken bij het ontwerpen van beoordelingstaken’.

Het cluster *Docent als coach* (n = 8) heeft een gemiddelde brugwaarde van 0.32 en is dus redelijk consistent. Het draait in dit cluster om het aanleren van reflectievaardigheden waarin de docent een begeleidende rol speelt. Het cluster bevat stellingen als ‘als student na te denken en in gesprek te gaan met zijn begeleider over zijn leren op een betekenisvol moment’, ‘beoordelen zien als een ontdekkingsreis voor de student’ en ‘de student stimuleren eerlijk en onbevangen naar zichzelf te kijken’.

Het cluster *Hulpmiddelen* (n = 4) heeft een gemiddelde brugwaarde van 0.17 en is dus zeer consistent. Het gaat hierbij om concrete manieren waarop duurzaam beoordelen vormgegeven zou kunnen worden. Het cluster bevat stellingen als ‘modeling examples gebruiken’, ‘het portfolio als basis te nemen’ en ‘de student de juiste woordenschat aan te leren’. Dit cluster ligt in het midden van de clustervisualisatie, wat betekent dat deze de meeste samenhang vertoont met de omliggende clusters.

Het cluster *Zelfsturing* (n = 18) heeft een gemiddelde brugwaarde van 0.09 en is daarmee het meest consistente cluster. Het gaat in dit cluster om het aanleren van vaardigheden en attitudes ter bevordering van eigenaarschap en zelfsturing van de student. Het bevat stellingen als ‘de student leren hoe zij het beste leren’, ‘studenten inzicht te geven in hun eigen professionele identiteit’, ‘de student eigenaarschap te geven over het leerproces’ en ‘studenten leren kritisch te zijn op eigen werk’.

Het cluster *Organisatiecultuur* (n = 12) heeft een gemiddelde brugwaarde van 0.25 en is dus redelijk consistent. Het gaat hierbij om het creëren van een duurzame mindset bij docenten en alle lagen van de organisatie. Het bevat stellingen als ‘toewijding te hebben op alle niveaus van de organisatie’, ‘een duidelijke visie op beoordelen hebben’, ‘collega’s de noodzaak hiervan te laten inzien’ en ‘de mate van duurzaamheid van kennis te erkennen’.

Het cluster *Voorwaarden voor implementatie* (n = 10) heeft een gemiddelde brugwaarde van 0.40 en is daarmee matig consistent. Het gaat hierbij om randvoorwaarden zoals tijd en ruimte voor

onderzoek om duurzaam beoordelen te kunnen implementeren in het onderwijs. Het cluster bevat stellingen als ‘de tijd krijgen en te nemen om samen te bouwen aan een werkbare methodiek’, ‘eerlijk en kritisch plannen te maken’, ‘blijven ontwikkelen na invoering’, en ‘duurzaam beoordelen geïntegreerd aan te bieden als logisch onderdeel van het praktijk onderwijs’.

Het cluster *Ipsatief beoordelen* ($n = 6$) heeft een gemiddelde brugwaarde van 0.77 wat aangeeft dat de stellingen in dit cluster niet erg eenduidig zijn. Inhoudelijk lijkt er inderdaad een tweedeling. Enerzijds zijn er stellingen met betrekking op beoordelingscriteria zoals ‘heldere beoordelingscriteria geven aan de student’ en ‘beoordelingspunten geven die duidelijk zijn voor zowel student als docent’. Anderzijds bevat het cluster stellingen met betrekking tot de mindset die van een docent verwacht wordt, zoals ‘flexibel te kunnen zijn in de mate waarin je verlangt dat de student eigenaarschap over zijn ontwikkeling neemt’ en ‘beseffen dat ieder student anders leert in een eigen tempo’. Toch zijn beide te verenigen in de term ipsatief beoordelen, waarbij heldere beoordelingscriteria voorwaardelijk zijn om groei ten opzichte van jezelf te kunnen meten.

Het cluster *Groeimindset* ($n = 12$) heeft een gemiddelde brugwaarde van 0.45 wat aangeeft dat de stellingen in het cluster matig consistent zijn. Dit cluster bevat stellingen die samenhangen met het ontwikkelen van een mindset waarbij de student het belang van reflectie en feedback inziet voor zijn persoonlijke leerproces. Het bevat stellingen zoals ‘reflecteren op het proces, de aanpak en de context’, ‘reflecteren voorafgaand aan beoordelen’, ‘terug te kijken en vooruit te kijken’, en ‘bewust te worden van het leerproces’.

3.2 Resultaten scoringsdata

De mate waarin de stellingen belangrijk en haalbaar werden gevonden is door 11 respondenten gescoord op een 5-punts Likertschaal. Alle 84 stellingen zijn door deze respondenten gescoord (100 %). Tabel 3 geeft een overzicht van de gemiddelde scores op belang en haalbaarheid per cluster en de bijbehorende statistieken.

Tabel 3

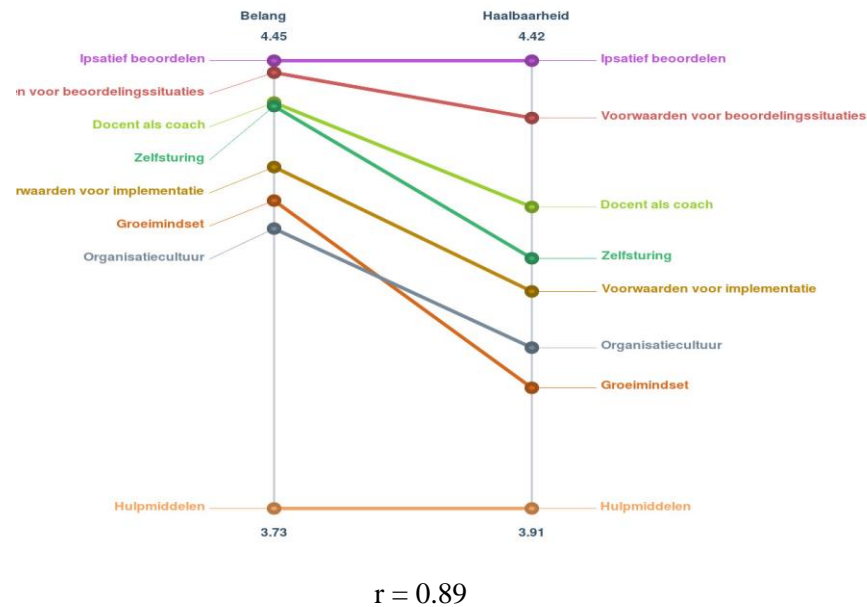
Acht clusters met gemiddelde scores van belang en haalbaarheid per cluster, variantie, minimum, maximum en mediaan.

<i>Cluster</i>	<i>Aantal stellingen</i>	<i>Belang/ Haalbaarheid</i>	<i>Gemiddelde score</i>	<i>Variantie</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mediaan</i>
1 Beoordelings-situaties	14	Belang	4.43 (0.37)	0.14	3.55	4.90	4.59
		Haalbaarheid	4.35 (0.28)	0.08	3.82	4.91	4.36
2 Docent als coach	8	Belang	4.38 (0.19)	0.04	4.18	4.63	4.32
		Haalbaarheid	4.25 (0.30)	0.09	3.82	4.64	4.18
3 Hulpmiddelen	4	Belang	3.73 (0.29)	0.08	3.45	4.09	3.68
		Haalbaarheid	3.91 (0.17)	0.03	3.72	4.10	3.90
4 Zelfsturing	18	Belang	4.38 (0.36)	0.13	3.33	4.72	4.45
		Haalbaarheid	4.19 (0.27)	0.07	3.64	4.72	4.18
5 Organisatie-cultuur	12	Belang	4.18 (0.38)	0.15	3.50	4.63	4.36
		Haalbaarheid	4.09 (0.32)	0.11	3.55	4.45	4.14
6 Voorwaarden implementatie	10	Belang	4.28 (0.17)	0.03	4.00	4.54	4.32
		Haalbaarheid	4.28 (0.32)	0.10	3.27	4.36	4.27
7 Ipsatief beoordelen	6	Belang	4.45 (0.16)	0.03	4.27	4.67	4.41
		Haalbaarheid	4.42 (0.26)	0.07	4.00	4.64	4.54
8 Groeimindset	12	Belang	4.20 (0.38)	0.15	3.63	4.81	4.32
		Haalbaarheid	4.22 (0.48)	0.24	2.90	4.64	4.18

Uit Tabel 3 valt op te maken dat alle clusters als relatief belangrijk worden ervaren ($M > 3.73$). Het cluster *Ipsatief beoordelen* scoort gemiddeld het hoogst op belang ($M = 4.45$). Ook de clusters *Groeimindset*, *Zelfsturing*, *Docent als coach* en *Voorwaarden voor beoordelingssituatie* scoren relatief hoog op belang. Het cluster *Hulpmiddelen* ($M = 3.73$) wordt als het minst belangrijk ervaren. Alle clusters worden ook als relatief haalbaar ervaren ($M > 3.91$). De clusters *Ipsatief beoordelen* ($M = 4.42$) en *Voorwaarden voor beoordelingssituatie* ($M = 4.35$), worden als haalbaarste clusters beschouwd. Het cluster *Hulpmiddelen* ($M = 3.91$) wordt als het minst haalbaar ervaren. Opvallend is dat het cluster *Groeimindset* de grootste variantie ($s^2 = 0.24$) vertoont op haalbaarheid met minimum (2.90) en maximum (4.64) scores, wat duidt op verschillen in mening over de haalbaarheid van de onderliggende items. Ook op belang is de variantie binnen dit cluster ($s^2 = 0.15$) relatief hoog. Op item-niveau wordt de stelling ‘de student zich veilig laten voelen’ het hoogst gescoord op belang ($M = 4.91$) en de stelling ‘studenten direct zelfstandig laten werken’ ($M = 3.3$) wordt als minst belangrijk gezien. De stelling ‘de student zich veilig laten voelen’ wordt als het meest haalbaar gezien ($M = 4,9$)

en de stelling ‘studenten zelf het eindniveau laten bepalen’ ($M = 2.9$) wordt als minst haalbaar gezien. Bijlage D geeft het volledige overzicht van de gemiddelde score per item op belang en haalbaarheid.

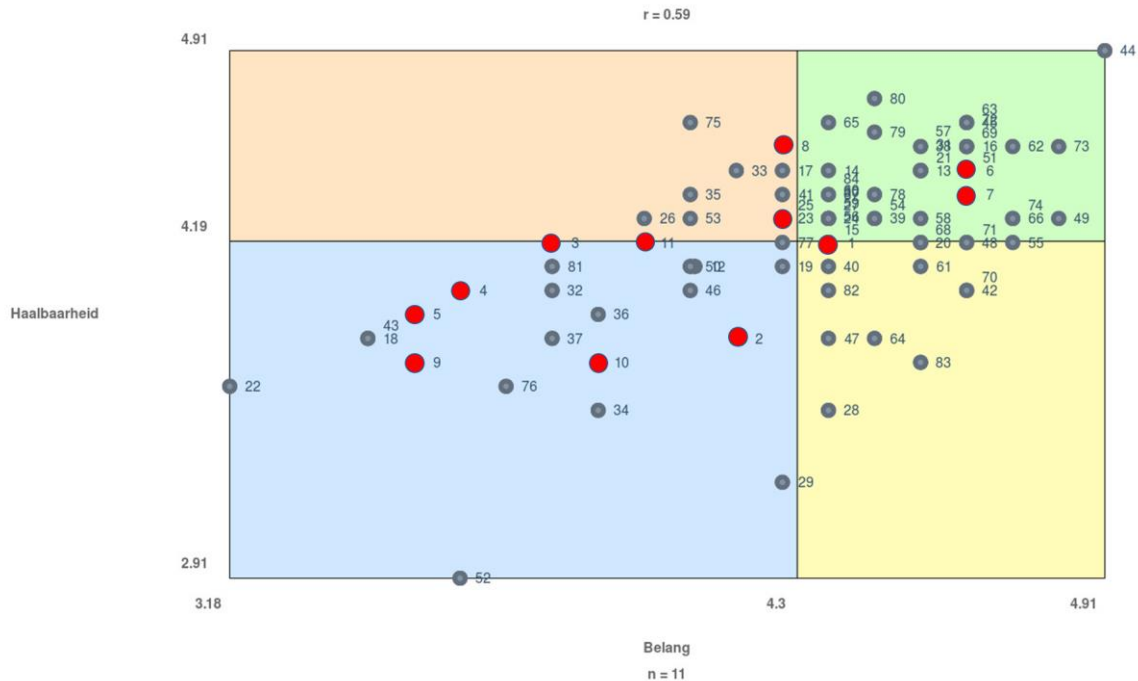
De pattern-match in Figuur 7 toont de vergelijking tussen belang en haalbaarheid op basis van de gemiddelde clusterwaarden.



Figuur 7. Pattern-match tussen belang en haalbaarheid en de correlatie.

Opvallend is dat het cluster *Ipsatief beoordelen* zowel het meest belangrijk als het meest haalbaar wordt gezien en het cluster *Hulpmiddelen* zowel het minst belangrijk als het minst haalbaar. De correlatie ($r = 0.89$) is erg hoog, wat betekent dat wanneer een cluster hoog op belang scoort, het waarschijnlijk ook hoog op haalbaarheid scoort. Dat het cluster *Ipsatief beoordelen* hoog scoort op zowel haalbaarheid als belang is te wijten aan twee stellingen in dit cluster namelijk ‘kritisch zijn op eigen handelen en dat van anderen studenten/docenten’ ($M = 4,6$) en ‘vanuit de docent gezien, beseffen dat iedere student anders leert en in een ander tempo’ ($M = 4,6$). Omdat er relatief weinig stellingen in dit cluster zitten, halen deze twee stellingen het clustergemiddelde enorm omhoog. Gezien het gebrek aan interne consistentie binnen dit cluster is het van belang om de hoge gemiddelde scoringswaarde van dit cluster te nuanceren en vooral naar de scores van de individuele stellingen in dit cluster te kijken.

De relatie tussen belang en haalbaarheid per stelling wordt in de Go-Zone (Figuur 8) weergegeven.



Figuur 8. Go-Zone met in rood gemarkeerd de theoretische ontwerpprincipes.

De groene zone ($n = 32$) bevat stellingen die hoog scoren op zowel belang als haalbaarheid en welke dus zeker meegenomen dienen te worden in vervolgacties ter bevordering van de implementatie van duurzaam beoordelen. Voorbeeld items uit de groene zone die het hoogst scoren zijn: ‘de student zich veilig te laten voelen’, en ‘tijd te nemen voor overleg en het gesprek aan gaan met de student’. Items in de blauwe zone ($n = 23$) zijn het laagst scoort op zowel belang als haalbaarheid. Stellingen in de blauwe zone die het laagst scoren zijn: ‘de student direct zelfstandig te laten werken’ en ‘studenten zelf het eindniveau te laten bepalen’. Opvallend is dat zeven van de twaalf ontwerpprincipes uit de literatuur over duurzaam beoordelen in de blauwe zone ($n = 23$) zijn beland (zie rode stippen) en dus door docenten als niet belangrijk en haalbaar worden gezien. Van de overige ontwerpprincipes uit de literatuur zijn er twee in de oranje zone, één in de gele zone en slechts twee stellingen in de groene zone beland. De theoretische ontwerpprincipes die wel als belangrijk en haalbaar worden gezien zijn: ‘dat de student leert zijn eigen prestatie te beoordelen op basis van zelfreflectie en externe feedback, ‘dat zowel de student als docent informatierijke feedback geeft en ontvangt’.

4. Conclusie en discussie

Het doel van dit ontwerpgerichte onderzoek was om ontwerpprincipes te formuleren ter bevordering van de duurzame beoordelingspraktijk van de participerende mbo-instelling. De centrale onderzoeksvraag was: Welke ontwerpprincipes ter bevordering van duurzaam beoordelen kunnen worden geformuleerd op basis van opvattingen van docenten en de wetenschappelijke literatuur? De cluster-indelingen en itemscores die door middel van GCM gegenereerd zijn bieden een basis: 1) waaruit ontwerpprincipes kunnen worden geformuleerd die een gereede kans hebben door docenten in de praktijk te worden geïmplementeerd en 2) waarmee de in de literatuur beschreven uitgangspunten van duurzaam beoordelen vanuit een praktijkperspectief kunnen worden gevalideerd en gecontextualiseerd.

Door middel van een brainstorm zijn 73 unieke stellingen over duurzaam beoordelen door docenten van een mbo-instellingen aangedragen. Aanvullend zijn 11 stellingen over duurzaam beoordelen uit de literatuur toegevoegd aan de totale statements-set. Zowel de opvattingen van docenten als de theoretische ontwerpprincipes zijn door de docenten gesorteerd en op basis van de clusteranalyse onderverdeeld in acht clusters: 1) *Voorwaarden voor beoordelingssituaties*, 2) *Docent als coach*, 3) *Hulpmiddelen*, 4) *Zelfsturing*, 5) *Organisatiecultuur*, 6) *Voorwaarden voor implementatie*, 7) *Ipsatief beoordelen*, 8) *Groeimindset*. De clusterlabels laten zien welke thema's kunnen worden onderscheiden vanuit het praktijkperspectief van de docenten. De theoretische ontwerpprincipes zijn door docenten gelijkmatig in de verschillende clusters onderverdeeld, waardoor in ieder cluster zowel docentenopvattingen als theoretische inzichten vertegenwoordigd zijn. Om antwoord te geven op de onderzoeksvraag is daarom gekozen om de ontwerpprincipes te formuleren op basis van de clusters, omdat dit leidt tot een synthese van opvattingen van docenten- en theoretische inzichten. Tot slot bieden de scoringsresultaten aanknopingspunten voor het prioriteren van ontwerpprincipes waar docenten in geloven en waar ze ook haalbare resultaten verwachten. Hierbij moet wel direct de kanttekening worden geplaatst dat met slechts 11 participanten deze resultaten niet voldoende representatief zijn voor de gehele organisatie.

De volgende ontwerpprincipes ter bevordering van duurzaam beoordelen kunnen worden geformuleerd op basis van de clusterthema's waarin opvattingen van docenten en wetenschappelijke literatuur elkaar complementeren:

- 1) Realiseer voor duurzaam beoordelen een veilige en authentieke beoordelingssituatie voorwaardelijk voor experimenteren en fouten maken door studenten.
- 2) Zorg bij duurzaam beoordelen dat de docent studenten coacht bij reflecteren op eigen leerprestatie en externe feedback.
- 3) Maak bij duurzaam beoordelen gebruik van hulpmiddelen zoals modeling examples en portfolio's.

- 4) Bevorder bij duurzaam beoordelen de zelfsturende vaardigheden van de student zodat deze tot een geïnformeerd oordeel kan komen over de eigen leerprestatie.
- 5) Zorg bij duurzaam beoordelen voor een organisatiecultuur met een gedeelde visie in alle lagen van de organisatie.
- 6) Zorg bij duurzaam beoordelen voor ruimte voor onderzoek en ontwikkeling in docenten-teams om tot een curriculum-aanpak te kunnen komen.
- 7) Zorg bij duurzaam beoordelen voor een ipsatief beoordelingsmodel waarbij student en docent aan de hand van heldere beoordelingscriteria groei ten opzichte van zichzelf kunnen meten.
- 8) Zorg voor een groei mindset bij student en docent zodat het belang van het ontwikkelen van zelfsturende vaardigheden ingezien wordt.

Deze acht ontwerpprincipes dienen als richtlijnen voor het bevorderen van een duurzame beoordelingspraktijk bij de participerende onderwijsinstelling en beschrijven voorwaarden voor docent- en studentgedrag, curriculumontwerp en organisatie. Om bij de stem van de deelnemers te blijven, is ervoor gekozen om ook de randvoorwaardelijke clusters als ontwerpprincipes te formuleren. Daardoor zijn niet alle uit de cluster map voortkomende ontwerpprincipes relevant als handvatten voor docenten om hun beoordelingspraktijk in te richten, maar worden ook randvoorwaardelijke ontwerpprincipes voor de bredere organisatie beschreven. Het oorspronkelijke doel van dit onderzoek om ontwerpprincipes te formuleren die een gerede kans hebben door docenten in de praktijk te worden geïmplementeerd moet dus breder worden geformuleerd. Dit onderzoek levert ontwerpprincipes op die niet allemaal direct in de invloedssfeer van docent zelf liggen, maar die docenten wel belangrijk vinden. Wat volgt is een inhoudelijke discussie met betrekking tot ieder individueel ontwerpprincipe.

Het eerste ontwerpprincipe is gebaseerd op de stellingen uit het cluster *Voorwaarden voor beoordelingssituaties*. De theoretische ontwerpprincipes die door docenten bij dit cluster ingedeeld zijn, zijn: 'authentieke en betekenisvolle beoordelingssituaties realiseren', 'de student betrekken in het ontwerpen van beoordelingstaken' en 'zowel de student als docent geeft en ontvangt informatierijke feedback'. Docenten dragen in dit cluster zelf stellingen aan waaruit blijkt dat een veilige beoordelingsomgeving, waarin de docent en student elkaar kennen en vertrouwen en waarin ruimte is om te experimenteren en fouten te maken, voorwaardelijk is voor een duurzame beoordelingspraktijk. De stelling 'de student zich veilig laten voelen' is tevens de hoogst scorende stelling op belang en haalbaarheid. Dat een veilige leeromgeving belangrijk is voor het realiseren van een duurzame beoordelingspraktijk is ook in de literatuur bekend. Een belangrijke voorwaarde voor het kunnen ontstaan voor feedback is een veilige leeromgeving (Govaerts et al., 2007; Liu & Carless, 2006; Van Diggelen, Beijgaard & den Brok, 2013). De docent heeft een belangrijke rol in het creëren van een veilige leeromgeving, waarin de student feedback en het maken van fouten ervaart als een instrument om prestaties te verbeteren (Fasté et al., 2012).

Het tweede ontwerpprincipe is gebaseerd op de stellingen uit het cluster *Docent als coach*. Het theoretische ontwerpprincipe: ‘de unieke student centraal stellen door middel van een ipsatief beoordelingsmodel’, is door docenten in dit cluster ingedeeld. Dit theoretische ontwerpprincipe sluit inhoudelijk minder goed aan bij de overige stellingen in het cluster wat ook bevestigd wordt door de hoge brugwaarde van dit principe. Op basis van de stellingen die door docenten zijn aangedragen, wordt vooral de begeleidende rol van docent als coach in dit cluster zichtbaar. Zij zien hun rol in een duurzame beoordelingspraktijk als begeleider die de student helpt op zijn ontdekkingsreis en hem helpt leren omgaan met externe feedback en kritische zelfreflectie. Dit ontwerpprincipe sluit aan bij het model van Fastré et al. (2012), waarin coaching als één van de drie methoden voor het bevorderen van de zelfsturende vaardigheden van de student wordt neergezet. In dit model wordt coaching beschreven als het helpen interpreteren en valideren van zelfbeoordelingen en peer-assessment. Het is daarom niet verwonderlijk dat er inhoudelijke overlap lijkt met het cluster *Zelfsturing* dat naast het cluster *Docent als coach* ligt in de cluster map. Beide clusters scoren ook hoog op belang. De coachende docent helpt de student om de zelfsturende vaardigheden eigen te maken. Opvallend is wel dat het gebruik van peer-assessment niet door docenten genoemd wordt in dit of één van de andere clusters. Dit sluit mogelijk aan bij de bevindingen van Davis en Neitzel (2011) dat docenten verantwoordelijkheid voor het beoordelingsproces veelal bij zichzelf leggen.

Het derde ontwerpprincipe komt voort uit het cluster *Hulpmiddelen*. Het theoretische ontwerpprincipe dat in dit cluster is ingedeeld is: ‘modeling examples gebruiken om de student te laten zien wat van ze verwacht wordt’. In het model van Fastré et al. (2012) is het gebruik van modeling examples één van de drie methodes, naast scaffolding en coaching, voor het bevorderen van een duurzame beoordelingspraktijk. Het cluster ligt in het midden van de cluster map, dichtbij het cluster *Zelfsturing* (waarin scaffolding genoemd wordt) en *Docent als coach*. De drie methodes uit het model van Fastré et al. zijn dus ieder in een eigen cluster verdeeld, maar hebben wel onderlinge verwantschap in de cluster map. Interessant is ook dat de onderzoekers een oproep doen tot verder onderzoek ter aanvulling van het model. Docenten zelf zien het gebruik van portfolio’s en het aanleren van een woordenschat om over duurzaam beoordelen te kunnen praten als aanvullingen op het gebruik van modeling examples. Opvallend is wel dat docenten het gebruik van hulpmiddelen als het minst belangrijk en haalbaar zien om een duurzamen beoordelingspraktijk te realiseren.

Het vierde ontwerpprincipe komt voort uit het cluster *Zelfsturing*. Het is het cluster met de laagste gemiddelde brugwaarde wat aangeeft dat er grote overeenstemming was tussen de participanten om deze stellingen bij elkaar te plaatsen. Het cluster scoort ook net als *Docent als coach* hoog op belang. Het gaat in dit cluster om het aanleren van vaardigheden en attitudes ter bevordering van eigenaarschap en zelfregulatie van de student. De theoretische ontwerpprincipes: ‘dat de student leert zijn eigen prestatie te beoordelen op basis van zelfreflectie en externe feedback’ en ‘een actieve rol en (toenemende mate van) eigenaarschap van de student in het gehele beoordelingsproces door middel van scaffolding te realiseren’, zijn door docenten in dit cluster ingedeeld. Stellingen die door docenten

zelf zijn aangedragen zijn ‘de student leren hoe zij het beste leren’, ‘studenten inzicht geven in hun eigen professionele identiteit’, ‘de student eigenaarschap geven over het leerproces’ en ‘studenten leren kritisch te zijn op eigen werk’. De stellingen hebben betrekking op het aanleren van diverse vaardigheden die studenten volgens de docenten nodig hebben om hun zelfsturende vaardigheden te vergroten. Hierbij worden vaardigheden als leren communiceren, leren leren, leren eigen leerdoelen te formuleren, leren reflecteren, leren feedback geven en erom te vragen, inzicht krijgen in eigen leerstijl en eigen professionele identiteit genoemd. Ook in dit cluster is overlap te vinden met uitgangspunten uit het onderzoek van Fastré et al. (2012). Zij zien de actieve houding van een student als één van de drie voorwaardelijke condities voor het ontwikkelen van duurzame beoordelingsvaardigheden. Fastré en collega’s zien omgaan met feedback als belangrijk onderdeel hiervan, met name het ontwikkelen van een geïnformeerd oordeel over je eigen leerprestatie door interne feedback (zelfreflectie) te vergelijken met externe feedback en daar vervolgacties uit te formuleren.

Het vijfde en zesde ontwerpprincipes komen voort uit de clusters *Organisatiecultuur* en *Voorwaarden voor implementatie*. Beide cluster liggen dicht bij elkaar in de clustermap en hebben inhoudelijke verwantschap. In het cluster *Organisatiecultuur* wordt het belang van een duurzame mindset genoemd dat bij docenten, maar ook in alle lagen van de organisatie aanwezig moet zijn. De theoretische ontwerpprincipes: formeel en non-formeel leren erkennen en de mate van duurzaamheid van kennis erkennen zijn in dit cluster gesorteerd. Bij het cluster *Voorwaarden van implementatie* worden praktische voorwaarden genoemd met betrekking tot tijd en ruimte, zoals tijd om te bouwen aan een werkbare methodiek, om te blijven ontwikkelen na invoering en om verder onderzoek te doen naar de rol van de docent en de student. Deze docentenopvattingen komen niet terug in de wetenschappelijke literatuur van duurzaam beoordelen, maar hebben wel raakvlakken met de literatuur omtrent formatief beoordelen. In het onderzoek van Schildkamp et al. (2014) worden voorwaarden voor de context genoemd voor effectieve formatieve toetsing, bestaande uit vijf aandachtsgebieden: 1) dat er een formatieve toetscultuur moet zijn gericht op samenwerking en continue verbetering 2) voorwaarden voor benodigde tijd en middelen 3) ruimte voor docentprofessionalisering 4) leiderschap en visie en 5) context buiten de school. Veel van de docentenstellingen uit de clusters *Organisatiecultuur* en *Voorwaarden voor implementatie* zijn herkenbaar in deze contextuele voorwaarden voor formatieve toetsing. Gezien de grote overlap tussen formatief en duurzaam beoordelen is het aannemelijk dat dezelfde voorwaarden ook relevant zijn voor de implementatie van duurzaam beoordelen.

Het zevende ontwerpprincipes komt voort uit het cluster *Ipsatief beoordelen*. Vanwege de hoge brugwaarde van dit cluster is het lastig om conclusie te verbinden aan dit cluster. Het theoretische ontwerpprincipes dat in dit cluster gesorteerd is; ‘heldere beoordelingscriteria te geven en de student te leren deze ook zelf te formuleren’. Inhoudelijk lijken er twee verschillende onderwerpen in het cluster aan bod te komen, wat mogelijk ook de hoge brugwaarde van het cluster verklaart. Enerzijds worden stellingen aangedragen die aansluiten bij het theoretische ontwerpprincipes, waarin de nadruk ligt op

het belang van duidelijke beoordelingscriteria. Anderzijds worden stellingen aangedragen met betrekking tot het flexibel kunnen omgaan met deze beoordelingscriteria, de focus op groei en het besef dat iedere student anders is en leert. Toch zijn deze stellingen inhoudelijk wel te verenigen in het concept ipsatief beoordelen, waarin groei ten opzichte van jezelf in plaats van anderen het uitgangspunt is. Sluijsmans (2010) benoemt dat heldere standaarden en criteria voorwaardelijk zijn voor een ipsatief beoordelingsmodel. Het theoretische ontwerpprincipes waarin het ipsatieve beoordelingsmodel centraal staat is vreemd genoeg niet bij dit cluster ingedeeld, maar bij het cluster *Docent als coach*. Een mogelijke verklaring hiervoor is dat de term ipsatief beoordelen nog onvoldoende bekend is onder docenten wat kan leiden tot verschil in interpretatie.

Niet alleen de term ipsatief beoordelen, maar duurzaam beoordelen als geheel lijkt een relatief complex onderwerp. Uit de Go-Zone blijkt bijvoorbeeld dat het merendeel van de theoretische ontwerpprincipes door docenten als niet belangrijk en niet haalbaar gescoord zijn. De Go-Zone lijkt hiermee te bevestigen dat docenten de theoretische ontwerpprincipes niet herkennen en deze dus ook niet als belangrijk en haalbaar bestempelen. Tijdens de terugkoppelingssessie werd dit ook door de 8 participanten bevestigd. Het lijkt erop dat de taal van de theoretisch ontwerpprincipes te abstract is en niet overeen komt met de taal die door de docenten zelf gesproken wordt. Dit ondersteunt het uitgangspunt van dit onderzoek dat de conceptuele aard van de bestaande wetenschappelijke literatuur omtrent duurzaam beoordelen, een vertaalslag naar de praktijk in de weg staat (Boud & Soler, 2016; Fastré, Van der Klink, Sluijsmans, & Merriënboer, 2012).

Het achtste en laatste ontwerpprincipes komt voort uit het cluster *Groei-mindset*. Het theoretische ontwerpprincipes dat in dit cluster gesorteerd is: ‘alle elementen van de complexe zelfsturende beoordelingsvaardigheid van de student te stimuleren, waardoor de student een geïnformeerd oordeel ontwikkelt’. Stellingen van docenten in dit cluster hebben vooral betrekking op het leerproces. Het gaat hierbij om het creëren van bewustzijn bij de student over zijn leerproces om te leren inschatten wat hij/zij wanneer nodig heeft. Het cluster ligt dicht bij het cluster *Zelfsturing*. Dit is inhoudelijk ook logisch, omdat het in het cluster *Groei-mindset* gaat om een bewustwording van het belang van feedback en zelfreflectie om je persoonlijke leerproces te bevorderen. Deze mindset is voorwaardelijk om de bijbehorende zelfsturende vaardigheden eigen te kunnen maken. Het ontwikkelen van een groei-mindset wordt door docenten als redelijk belangrijk gezien, maar veel minder haalbaar. Opvallend is wel dat de meningen over de haalbaarheid sterk uiteenlopen, wat blijkt uit de relatief grote verschillen tussen de minimum en maximum scores in vergelijking met de andere clusters.

Zoals eerder al werd aangegeven is het lastig om op basis van de scoringsdata van 11 participanten uitspraak te doen over het belang en de haalbaarheid van de ontwerpprincipes. Een andere beperking met betrekking tot de generaliseerbaarheid van de onderzoeksresultaten is de contextgebonden onderzoekopzet. De dataverzameling vond plaats binnen één instelling voor het beroepsonderwijs. Dit is voor het ontwikkelen van ontwerpprincipes binnen deze organisatie een voordeel, maar voor de generalisatie van de resultaten is deze opzet een beperking. Ook bij de

generaliseerbaarheid binnen de organisatie zijn kanttekeningen te plaatsen. Niet ieder onderwijsteam is in het onderzoek vertegenwoordigd en er is een overrepresentatie van andere onderwijsteams, waardoor de resultaten geen representatief beeld vormen voor de gehele organisatie. De kracht van GCM in het zoveel mogelijk betrekken van de onderzoeksdoelgroep is tevens ook een beperking van deze onderzoeksmethode. GCM vraagt veel tijd van de deelnemers, met name de sorteer en scoringsfase en dit verklaart wellicht deels de afnemende participantenaantallen bij iedere volgende fase van het onderzoek.

Een tweede beperking is de interne consistentie van een drietal clusters die matig tot slecht zijn te noemen. De zorgvuldigheid waarmee de sorteeractiviteiten zijn uitgevoerd, bepaalt in grote mate de kwaliteit van de clusteranalyse en de clusteroplossing. Het is daarom van belang dat de instructies die de participanten ontvangen voor de sorteeractiviteit helder zijn en dat zij ook echt de tijd nemen om deze activiteit zorgvuldig uit te voeren. In het huidige onderzoek vond de instructie enkel via de online tool plaats, waardoor participanten niet om verduidelijking konden vragen. Het is wellicht verstandig deze fase tijdens een gepland moment te laten plaats vinden, zodat er een mondelinge instructie en toelichting kan worden gegeven. Het nadeel hiervan is dan wel dat participanten niet in hun eigen tijd kunnen deelnemen, wat weer een negatief effect op participantenaantallen kan hebben.

Tot slot zijn er veel momenten in het proces van GCM waarbij de onderzoeker keuzes moet maken op basis van eigen inzichten. Om dit zo objectief mogelijk te doen is een tweede en soms derde beoordelaar bij deze keuzes betrokken geweest, maar het is lastig om elke vorm van subjectiviteit geheel te elimineren. Met name bij het formuleren van de ontwerpprincipes zijn keuzes gemaakt om de individuele stellingen tot één ontwerpprincipe te synthetiseren, waarbij het onmogelijk is om alle onderliggende items in een generiek principe samen te vatten. Het is daarom aan te raden om de onderliggende individuele stellingen niet uit het oog te verliezen.

Ondanks de beperkingen heeft die onderzoek ontwerpprincipes opgeleverd die kunnen worden gebruikt om interventies in te richten waarvan docenten het belang ook daadwerkelijk inzien. Aanbeveling voor verder onderzoek is om door middel van een interventiestudie te kijken naar de effecten van interventies gebaseerd op deze ontwerpprincipes. Vervolgonderzoek zou moeten uitwijzen of de gehanteerde onderzoeksmethode inderdaad geschikt is voor het creëren van draagvlak gevolgd door succesvolle implementatie. Een tweede aanbeveling is om vergelijkbare onderzoeken uit te voeren op andere mbo-instellingen of andere onderwijsgroepen waar een behoefte leeft aan verdere implementatie van duurzaam beoordelen. De resultaten van onderzoeken binnen andere mbo's kunnen worden geaggregeerd, wat kan bijdragen aan het inventariseren van thema's waarop een landelijke onderzoeksagenda kan worden gebaseerd. Tot slot benadert het huidige onderzoek het concept duurzaam beoordelen vanuit de docent. Een waardevolle toevoeging voor vervolgonderzoek is om ook de opvattingen van studenten in kaart te brengen.

Dit onderzoek is één van de eerste pogingen om vanuit een praktijkperspectief te kijken naar duurzaam beoordelen, binnen de relatief ongebruikelijke context van het mbo. In dit onderzoek is

GCM gebruikt voor het ontwikkelen van conceptuele kaders vanuit een praktijkperspectief wat bijdraagt aan het valideren en contextualiseren, maar ook het verfijnen en concretiseren van de in de theorie beschreven uitgangspunten omtrent duurzaam beoordelen. De bevindingen van dit onderzoek ondersteunen het probleem dat de conceptuele aard van de bestaande wetenschappelijke literatuur een vertaalslag naar de praktijk in de weg staat en kunnen bijdragen aan het dichten van deze kloof.

Kennis en inzicht in opvattingen van docenten over duurzaam beoordelen is bruikbaar voor veel praktijkinstellingen die worstelen met inrichten van hun beoordelingspraktijken. Bestaande praktijken sluiten vaak onvoldoende aan bij de ambitie om studenten voor te bereiden op een leven lang leren. De ontwerpprincipes die op basis van de cluster-indelingen en scoringsdata van dit onderzoek gegenereerd zijn, kunnen dienen als handvatten voor andere onderwijsinstellingen en docenten, voor het opzetten van implementatie acties voor het realiseren van een duurzame(re) beoordelingspraktijk.

Referenties

- Assessment Reform Group. (2002). Assessment for learning 10 principles: research-based principles to guide classroom practice. *Assessment Reform Group*, 1–3.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational Leadership*, 57(8), 28–33.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5, 7–74. <https://doi.org/10.1080/0969595980050102>
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167. <https://doi.org/10.1080/713695728>
- Boud, D. (2010). Assessment for developing practice. In *Education for Future Practice*, edited by J. Higgs, D. Fish, I. Goulter, S. Loftus, J.A. Reid, and F. Trede, 251–262. Rotterdam: Sense.
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413. <https://doi.org/https://doi.org/10.1080/02602938.2015.1018133>
- Brand-Gruwel, S., Bos, N., & Van der Graaf, A. (2019). Het vergroten van studiesucces in het hoger onderwijs: Belang van overtuigingen van docenten. *Pedagogische Studiën*, 96, 1–14.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice*, 11, 301–318. <https://doi.org/10.1080/0969594042000304609>
- Calderhead, J. (1996). Teachers: Beliefs and knowledge. In *Handbook of educational psychology*. (pp. 709–725). London: Prentice Hall International.
- Concept Systems Incorporated. (2017). *The concept system global max software guide: Group decisions for the 21st century* [PDF file]. Retrieved from <https://www.conceptsystems.com/content/uploads/2015/09/CS-Global-Max-Software-Guide.pdf>
- Creswell, J. W. (2014). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Groot-Britannië, Harlow: Pearson.
- Davis, D. S., & Neitzel, C. (2011). A self-regulated learning perspective on middle grades classroom assessment. *Journal of Educational Research*, 104(3), 202–215. <https://doi.org/10.1080/0022067100369014>
- Davison, M. L. (1983). *Multidimensional scaling*. New York, NY: John Wiley & Sons.
- Everitt, B. (1980). *Cluster analysis* (2nd ed.). New York, NY: Halsted Press.
- Fastré, G., Van der Klink, M., Sluijsmans, D., & Merriënboer, J. J. G. Van. (2012). Towards an integrated model for developing sustainable assessment skills. *Assessment & Evaluation in Higher Education*, 1–20. <https://doi.org/10.1080/02602938.2012.674484>

- Govaerts, M. J. B., Van der Vleuten, C. P. M., Schuwirth, L. W. T., and Muijtjens, A. M. M. (2007). Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Advances in Health Sciences Education* 12(2), 239–60. <https://doi.org/10.1007/s10459-006-9043-1>
- Gulikers, J., & Baartman, L. (2017). *Doelgericht professionaliseren: Formatieve toetspraktijken met effect! Wat DOET de docent in de klas?* Wageningen: Wageningen University & Research.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hawley, W. D., & Valli, L. (1999). The essentials of effective professional development. In *Teaching as the learning profession: Handbook of policy and practice* (pp. 127–179). San Francisco: Jossey-Bass.
- Jackson, K. M., & Trochim, W. M. K. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods*, 5(4), 307–336. <https://doi.org/10.1177/109442802237114>
- Kane, M., & Trochim, W. M. K. (2007). *Concept mapping for planning and evaluations*. Thousand Oaks, CA: Sage Publications.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.
- Liu, N., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279–90. <https://doi.org/10.1080/13562510600680582>
- Onderwijsraad. (2016). *Vakmanschap voortdurend in beweging: Verbeter de aansluiting tussen mbo, arbeidsmarkt en een leven lang leren* [PDF file]. Retrieved from <https://www.onderwijsraad.nl/publicaties/adviezen/2016/10/13/vakmanschap-voortdurend-in-beweging>.
- Rosas, S. R., & Kane, M. (2012). Quality and rigor of the concept mapping methodology: A pooled study analysis. *Evaluation and Program Planning*, 35(2), 236–245. <https://doi.org/10.1016/j.evalprogplan.2011.10.003>
- Rosas, S. R., & Kane, M. (2018). *Conversations about group concept mapping: Applications, examples, and enhancements*. Los Angeles: Sage Publications, Inc.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144. <https://doi.org/10.1007/BF00117714>
- Schildkamp, K., Heitink, M., Van Der Kleij, F., Hoogland, I., Dijkstra, A., Kippers, W., & Veldkamp, B. (2014). *Voorwaarden voor effectieve formatieve toetsing: Een praktische review*. Den Haag.
- Sluijsmans, D. (2008). *Betrokken bij beoordelen (lectorale rede)*. Nijmegen: Hogeschool van Arnhem en Nijmegen.
- Sluijsmans, D. (2010). Beoordelen met het oog op de toekomst. *Examens*, 1(5), 5–10. <https://doi.org/10.1007/s12494-010-0042-5>
- Sluijsmans, D., Joosten-ten Brinke, D., & Van der Vleuten, C. (2013). *Toetsen met leerwaarde: Een reviewstudie naar de effectieve kenmerken van formatief toetsen*. Den Haag: NWO.

- Thomas, L., Deaudelin, C., Desjardins, J., & Dezutter, O. (2011). Elementary teachers' formative evaluation practices in an era of curricular reform in Quebec, Canada. *Assessment in Education: Principles, Policy and Practice*, 18(4), 381–398. <https://doi.org/10.1080/0969594X.2011.590793>
- Thomas, M. (2012). Teachers' beliefs about classroom assessment and their selection of classroom assessment strategies. *Journal of Research and Reflections in Education*, 6(2), 103–112.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). Reston, VA: National Council of Teachers of Mathematics.
- Thurlings, M., & Van Diggelen, M. (2019). Perceptions of practical knowledge of learning and feedback among academic teachers. *European Journal of Engineering Education*, 1–22. <https://doi.org/10.1080/03043797.2019.1677559>
- Trigwell, K., Prosser, M., & Taylor, P. (1994). Qualitative differences in approaches to teaching first year university science. *Higher Education*, 27, 75–84. <https://doi.org/10.1007/BF01383761>
- Trochim, W. (1989). Concept mapping: Soft science or hard art? *Evaluation and Program Planning*, 12(1), 87-110. [https://doi.org/10.1016/0149-7189\(89\)90027-X](https://doi.org/10.1016/0149-7189(89)90027-X)
- Trochim, W., & McLinden, D. (2017). Introduction to a special issue on concept mapping. *Evaluation and program planning*, 60, 166-175. <https://doi.org/10.1016/j.evalprogplan.2016.10.006>
- Van Berkel, H., Bax, H., & Joosten-ten Brinke, D. (2017). Toetsen in het hoger onderwijs. In *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.
- Van Diggelen, M. R., Brok, den, P. J., & Beijgaard, D. (2013). Teachers' use of a self-assessment procedure: the role of criteria, standards, feedback and reflection. *Teachers and Teaching*, 19(2), 115-134. <https://doi.org/10.1080/13540602.2013.741834>
- Verloop, N., Van Driel, J., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research*, 35, 441–461. [https://doi.org/10.1016/S0883-0355\(02\)00003-4](https://doi.org/10.1016/S0883-0355(02)00003-4)

Bijlagen

Bijlage A

Informatiebrief en toestemmingsverklaring

Bijlage B

Totale statements-set op basis van de brainstorm en de wetenschappelijke literatuur

Bijlage C

Clusterrapport sorteerddata

Bijlage D

Go-Zone rapport scoringsdata

Bijlage A

Informatiebrief en toestemmingsverklaring

Beste collega,

Ik wil je vragen mee te doen aan een wetenschappelijk onderzoek. Meedoen is vrijwillig. Om mee te doen heb ik je schriftelijke toestemming nodig. Voordat je beslist of je wilt meedoen aan dit onderzoek, krijg je uitleg over wat het onderzoek inhoudt. Lees deze informatie rustig door en vraag de onderzoeker (Sanne Keetels) uitleg als je vragen hebt.

1. Doel van het onderzoek

Het doel van dit onderzoek is om praktijk- en theorie-geïnformeerde ontwerpprincipes te formuleren ter bevordering van een duurzame beoordelingspraktijk.

2. Achtergrond van het onderzoek

Mbo-instellingen hebben de opdracht om studenten voor te bereiden op een leven lang leren, zodat zij niet alleen tijdens de opleiding, maar ook daarna succesvol richting kunnen geven aan hun eigen (professionele) ontwikkeling. Dit vraagt echter om een andere manier van beoordelen en begeleiden, dat meer gericht is op het ontwikkelen van de zelfsturende vaardigheden van een student. Uit onderzoek blijkt dat het lastig is om een nieuwe beoordelingspraktijk in de klas te realiseren en dat opvattingen van docenten over beoordelen sterke invloed heeft op de gekozen beoordelingsstrategieën in de klas.

3. Wat meedoen inhoudt en wat er van je verwacht wordt

Het onderzoek bestaat uit meerdere fasen. Je bijdrage aan alle fasen van het onderzoek is zeer welkom, maar je kan zelf beslissen of/en aan welke (één of meerdere) fasen van het onderzoek je wilt deelnemen. Er worden geen audiovisuele opnames gemaakt tijdens het onderzoek.

Fase 1 - De brainstormfase. Voor deze fase van het onderzoek ontvang je per mail een link naar een online tool waar je gevraagd wordt stellingen te noteren met betrekking tot je opvattingen over duurzaam beoordelen. Je hebt hier ongeveer 15 minuten voor nodig.

Fase 2 - De sorteerfase. Voor deze fase van het onderzoek wordt je gevraagd de stellingen die uit de brainstormfase en literatuurstudie zijn voortgekomen, te sorteren in clusters via een online tool. Je ontvangt een link naar deze tool via je email. Je hebt hier ongeveer 30 minuten voor nodig.

Fase 3: De scoringsfase. Voor deze fase van het onderzoek word je gevraagd de stellingen uit de brainstormfase te scoren op een vijfpunts-schaal via een online tool. Je ontvangt een link naar deze tool via je email. Je hebt hier ongeveer 15 minuten voor nodig.

Fase 4: De terugkoppelingssessie. Tijdens deze face-to-face sessie worden de voorlopige resultaten van het onderzoek gepresenteerd en wordt je gevraagd input te geven op de interpretatie van de resultaten. Deze sessie duurt 1 uur.

4. Mogelijke voor- en nadelen

Er zijn geen voor- of nadelen bekend als gevolg van deelname aan het onderzoek.

5. Als je niet wilt meedoen of wilt stoppen met het onderzoek

Je beslist zelf of je meedoet aan het onderzoek. Deelname is vrijwillig. Als je niet wilt deelnemen heeft dat geen nadelige gevolgen. Als je wel meedoet, kunt je je altijd bedenken en toch stoppen, ook tijdens het onderzoek. Je hoeft niet te zeggen waarom je stopt. De gegevens die tot dat moment zijn verzameld, mogen worden gebruikt voor het onderzoek.

6. Einde van het onderzoek

Na het verwerken van alle gegevens word je uitgenodigd voor een terugkoppelingssessie over de uitkomsten van het onderzoek. De uitkomsten van het onderzoeken zullen ook bekend worden gemaakt via de interne communicatiepagina van de onderwijsinstelling.

7. Gebruik en bewaren van je gegevens

Voor dit onderzoek worden er persoonsgegevens verzameld, gebruikt en bewaard. Je wordt gevraagd naar je naam, contactgegevens, leeftijd, geslacht, het onderwijsteam en welke rol(len) je binnen de organisatie vertegenwoordigd. Het verzamelen, gebruiken en bewaren van deze gegevens is nodig om de vragen die in dit onderzoek worden gesteld te kunnen beantwoorden. De uitkomsten van het onderzoek zullen worden gedeeld met collega's. De gegevens die worden gedeeld bevatten geen informatie die tot jou te herleiden is. Ook in rapporten en publicaties over het onderzoek zijn de gegevens niet tot jou te herleiden.

Vertrouwelijkheid van je gegevens.

Om je privacy te beschermen krijgen je gegevens een code. Je naam en andere gegevens die je direct kunnen identificeren worden daarbij weggelaten. Je gegevens worden op deze wijze versleuteld. De sleutel van de code blijft veilig opgeborgen, binnen de Open Universiteit. Alleen de onderzoeker en de OU begeleider Migchiël van Diggelen hebben toegang tot de niet-versleutelde informatie.

Toegang tot je gegevens voor controle.

Om te kunnen beoordelen of het onderzoek op een betrouwbare wijze is uitgevoerd, kunnen leden van een visitatiecommissie inzage krijgen in de niet-versleutelde informatie.

Bewaartermijn gegevens

Je gegevens moeten 10 jaar worden bewaard door de Open Universiteit.

Meer informatie over je rechten bij verwerking van gegevens

Voor algemene informatie over uw rechten bij verwerking van je persoonsgegevens kun je de website van de [Autoriteit Persoonsgegevens](#) raadplegen. De privacy disclaimer van de Open Universiteit vindt u via www.ou.nl/privacy.

8. Heb je vragen?

Bij vragen kun je contact opnemen met Sanne Keetels (sanne.keetels@ou.nl) of met de OU begeleider Migchiël van Diggelen (michiël.vandiggelen@ou.nl).

9. Ondertekening toestemmingsformulier

Wanneer je voldoende bedenktijd hebt gehad, word je gevraagd te beslissen over deelname aan dit onderzoek. Door je schriftelijke toestemming geeft je aan dat je de informatie hebt begrepen en instemt met deelname aan het onderzoek. Je kunt dit toestemmingsformulier (zie wervingsbericht op portaal) digitaal ondertekenen en **voor 4 september** terugsturen naar de onderzoeker (s.keetels@sintlucas.nl).

Toestemmingsverklaring

voor deelname aan het wetenschappelijk onderzoek:

Implementatie van duurzaam beoordelen: een GCM studie voor praktijk- en theoriegeïnformeerde ontwerpprincipes voor het middelbare beroepsonderwijs.

☐ Ik geef toestemming om de tijdens deze studie verzamelde data te gebruiken voor wetenschappelijke onderzoek.

☐ Ik heb de informatiebrief gelezen/ontvangen met betrekking tot dit onderzoek en ik heb de mogelijkheid gehad om vragen te stellen over onduidelijkheden aan de onderzoeker.

☐ Ik begrijp dat alle informatie die ik lever voor dit onderzoek op een veilige manier wordt verzameld, anoniem wordt gepubliceerd en niet naar mij terug te herleiden is.

☐ Ik begrijp dat ik ten alle tijden kan stoppen met deelname aan het onderzoek zonder reden van opgaaf.

☐ De data wordt voor een periode van 10 jaar bewaard volgens de VSNU regels.

Ik ben bereid deel te nemen aan de volgende fasen van het onderzoek:

☐ De brainstormsessie voor het genereren van stellingen met betrekking tot duurzaam beoordelen via een online tool.

☐ De sorteerfase van het onderzoek waarbij ik de gegenereerde stellingen uit de brainstormfase sorteer in clusters via een online tool.

☐ De scoringsfase van het onderzoek waarbij ik de gegenereerde stellingen uit de brainstormfase score op een vijfpuntsschaal via een online tool.

☐ De terugkoppelingssessie waarbij ik input kan geven op de resultaten van het onderzoek.

Als je de bovengenoemde punten gelezen hebt en op de punten instemt door de cirkels te vervangen met kruisjes, kunt je dit toestemmingsformulier digitaal ondertekenen door je naam, geboortedatum en dag van vandaag in te vullen. Je kunt het formulier per mail terug zenden naar sanne.keetels@ou.nl.

Naam:

Geboortedatum:

Onderwijsteam:

Datum:

Bijlage B

Totale statements-set met stellingen uit de literatuur en de brainstorm

Focus prompt: Voor het realiseren van een duurzame beoordelingspraktijk in het dagelijks onderwijs waarmee we de student helpen voorbereiden op een leven lang leren is het van belang om...

ID	Stelling
L1	authentieke en betekenisvolle beoordelingssituaties te realiseren.
L2	de unieke student centraal te stellen door middel van een ipsatief beoordelingsmodel (groei ten opzichte van jezelf i.p.v. anderen)
L3	formeel en non-formeel leren te erkennen.
L4	de mate van duurzaamheid van kennis te erkennen.
L5	de student te betrekken bij het ontwerpen van beoordelingstaken.
L6	dat de student leert zijn eigen prestatie te beoordelen op basis van zelfreflectie en externe feedback.
L7	dat zowel student als docent informatierijke feedback geeft en ontvangt.
L8	heldere beoordelingscriteria te geven en de student te leren deze ook zelf te formuleren.
L9	modeling examples te gebruiken om de student te laten zien wat van ze verwacht wordt.
L10	alle elementen van de complexe zelfsturende beoordelingsvaardigheid van de student te stimuleren, waardoor de student een geïnformeerd oordeel ontwikkelt.
L11	een actieve rol en (toenemende mate van) eigenaarschap van de student in het gehele beoordelingsproces door middel van scaffolding te realiseren.
B12	te werken aan self-efficacy
B13	dat zowel de student als docent het ingeleverde werk beoordelen.
B14	tools paraat te hebben die bij onze mbo-studenten passen om het leerproces te bespreken, inzichtelijk te maken, te visualiseren en te plannen.
B15	taal paraat te hebben die bij onze mbo-studenten past om het leerproces te bespreken, inzichtelijk te maken, te visualiseren en te plannen.

- B16 het voor de student praktisch te maken.
- B17 het voor de student uitdagend te maken.
- B18 een andere term dan "beoordeling" te gebruiken.
- B19 beoordelen te zien als een ontdekkingsreis voor de student, samen met zijn begeleider waarin er geen goed of fout is.
- B20 te reflecteren op het proces, de aanpak, de context en allerlei verschillende factoren die bepalen in hoeverre iets goed of minder goed gelukt is.
- B21 te reflecteren voorafgaand aan beoordelen.
- B22 studenten direct zelfstandig te leren werken.
- B23 studenten te leren reflecteren op elkaars werk.
- B24 studenten te leren hoe je opbouwend communiceert.
- B25/ L12 het (duurzaam beoordelen) geïntegreerd aan te bieden als logisch onderdeel van de "practice"/"het creatieve proces".
- B26 er alert op te zijn "het" (duurzaam beoordelen) niet te theoretisch of analytisch aan te pakken.
- B27 in te zien hoe moeilijk reflectie en zelfbeoordeling is (de student is een mbo student) en dus is er opbouw (scaffolding) nodig.
- B28 de tijd te krijgen en te nemen om samen (als team) te leren en onderzoeken.
- B29 de tijd te krijgen en te nemen om samen (als team) te bouwen aan een werkbare methodiek.
- B30 te blijven ontwikkelen na invoering.
- B31 de begeleiding daar op af te stemmen.
- B32 bijvoorbeeld de feedbackcyclus in te zetten waarmee goed afgestemd kan worden op de behoefte van de student op ieder moment in het leerproces/project.
- B33 studenten te bevragen hoe zij dit voor zich zien.
- B34 toewijding te hebben op alle niveaus (docent --> leiding --> CvB) en het echt als SintLucas omarmen en uitdragen.
- B35 eerlijk en kritisch plannen te maken, rekening houdend met wat haalbaar is in duurzaam beoordelen in tijd en ruimte die we hebben.

- B36 studenten te leren hoe zij het beste leren binnen welke context.
- B37 studenten met verschillende typen portfolio's te laten werken.
- B38 leren centraal te stellen.
- B39 een duidelijke visie te hebben op beoordelen.
- B40 ruimte te geven aan collega's om te leren hoe zij dit het beste kunnen doen.
- B41 vertrouwen te hebben in de goede wil van de studenten.
- B42 collega's de noodzaak hiervan te laten inzien.
- B43 te voorkomen dat studenten zich blind staren op wat we stellen dat 'het proces' is.
- B44 de student zich veilig te laten voelen.
- B45 de student kennis te laten maken met opbouwende feedback.
- B46 de student de juiste woordenschat aan te leren.
- B47 de student te leren eigen leerdoelen en leerweg te bepalen binnen de kaders.
- B48 studenten ruimte/veiligheid te bieden voor het maken van fouten en verbeteren (na de beoordeling) zonder vertraging op te lopen.
- B49 tijd te nemen voor overleg en het gesprek aangaan met de student.
- B50 te beoordelen en reflecteren met een team.
- B51 de student te kennen.
- B52 studenten zelf het eindniveau laten bepalen.
- B53 te onderzoeken wat en in welke mate van een mbo-student verwacht kan worden binnen het aannemen van een actieve leerhouding/tonen van eigenaarschap.
- B54 formatieve toetsing een plek te geven binnen ons onderwijs.
- B55 de huidige student een andere beroepshouding aan te leren waarbij zij niet alleen komen 'halen' maar zelf een actievere leerhouding aannemen.
- B56 studenten inzicht te geven in hun eigen professionele identiteit.
- B57 studenten te leren reflecteren en formuleren van eigen leerdoelen.
- B58 de rol van de docent binnen het dagelijkse onderwijs eens goed te bekijken/onderzoeken.

- B59 een goede "mindset" aan het begin van de opleiding te bewerkstelligen (ook bij docenten); eigenaarschap over leren als logische competentie die je in de praktijk nodig hebt.
- B60 flexibel te kunnen zijn in de mate waarin je verlangt dat de student eigenaarschap over zijn/haar ontwikkelproces neemt en de verhouding tussen keuze en vaststaand programma daarop aanpast.
- B61 de student te stimuleren eerlijk en onbevangen naar zichzelf en zijn omgeving te kijken.
- B62 studenten te leren hoe ze van feedback, beoordelingen en reflecties kunnen leren.
- B63 studenten houvast te bieden bij het reflecteren en vragen om feedback.
- B64 de student echt eigenaarschap te geven over zijn leerproces, zowel voor de inhoud als de uitvoering.
- B65 te focussen op groei.
- B66 studenten te leren kritisch te zijn op eigen werk: zelfreflectie.
- B67 nieuwsgierigheid te prikkelen.
- B68 de student zich eigenaar en verantwoordelijk te laten voelen voor de 'beoordeling'.
- B69 de student niet af te rekenen op wat niet goed gaat, maar te letten op wat goed ging en wat beter kan.
- B70 kritisch te zijn op eigen handelen en dat van andere studenten/docenten.
- B71 te leren kritische en constructieve feedback te geven en te leren daarom te vragen.
- B72 vanuit de docent gezien, te beseffen dat iedere student anders leert en in een ander tempo.
- B73 te leren van 'fouten', accepteren dat in een leerproces een 'fout' je verder kan helpen en als gevolg daarvan durven experimenteren.
- B74 zicht te krijgen in eigen leerproces/leerstijl.
- B75 terug te kijken en vooruit te kijken; hoe ging ik een jaar geleden met deze materie om? hoe doe ik dat nu? en wat zou ik willen bereiken volgend jaar?
- B76 beoordelen te zien als waarderen van jouw inspanning voor jezelf en voor anderen (de docent, de klant, etc.).
- B77 bewust te zijn/worden (student en docent) van het leerproces, wat is van belang in deze fase van je leerproces.

- B78 een student onder begeleiding van een docent te helpen groeien in het kunnen reflecteren, het nemen van verantwoordelijkheid en in het vooruitkijken naar de (nabije) toekomst.
- B79 beoordelingspunten aan te geven die duidelijk zijn voor zowel student als docent.
- B80 de student te helpen bij het nadenken over haalbare leerdoelen.
- B81 het portfolio als basis te nemen waar de student zijn ontwikkelpunten op bij stelt naar gelang deze al aansluit bij het KD, ambities en dromen.
- B82 als student na te denken en in gesprek te gaan met zijn begeleider over zijn leren op een betekenisvol moment.
- B83 niet uit te gaan van een vaste route. De student zal naar aanleiding van ervaringen (vakinhoudelijk, maatschappelijk, persoonlijk) steeds weer zijn leerdoelen bijstellen.
- B84 dat de student de bedoeling begrijpt van een duurzame "beoordeling" voor zijn persoonlijk leerproces.

Bijlage C

Clustersrapport sorteerd data met brugwaarden en statistieken

Dikgedrukte stellingen komen uit de literatuur.

Cluster 1: voorwaarden voor beoordelingssituaties	0.37
1 authentieke en betekenisvolle beoordelingssituaties te realiseren.	0.35
5 de student te betrekken bij het ontwerpen van beoordelingstaken.	0.32
7 dat zowel student als docent informatierijke feedback geeft en ontvangt.	0.35
16 het voor de student praktisch te maken.	0.32
17 het voor de student uitdagend te maken.	0.47
37 studenten met verschillende typen portfolio's te laten werken.	0.39
38 leren centraal te stellen.	0.36
41 vertrouwen te hebben in de goede wil van de studenten.	0.31
44 de student zich veilig te laten voelen.	0.28
48 studenten ruimte/veiligheid te bieden voor het maken van fouten en verbeteren (na de beoordeling) zonder vertraging op te lopen.	0.39
49 tijd te nemen voor overleg en het gesprek aangaan met de student.	0.35
51 de student te kennen.	0.45
67 nieuwsgierigheid te prikkelen.	0.53
69 de student niet af te rekenen op wat niet goed gaat, maar te letten op wat goed ging en wat beter kan.	0.39

Aantal	SD	Variantie	Min.	Max.	Gem.	Mediaan
14	0.06	0.004	0.28	0.53	0.38	0.36

Cluster 2: Docent als coach	0.32
2 de unieke student centraal te stellen door middel van een ipsatief beoordelingsmodel (groei ten opzichte van jezelf i.p.v. anderen)	0.35
19 beoordelen te zien als een ontdekkingsreis voor de student, samen met zijn begeleider waarin er geen goed of fout is.	0.3
23 studenten te leren reflecteren op elkaars werk.	0.15
33 studenten te bevragen hoe zij dit voor zich zien.	0.28
45 de student kennis te laten maken met opbouwende feedback.	0.39
61 de student te stimuleren eerlijk en onbevangen naar zichzelf en zijn omgeving te kijken.	0.27
63 studenten houvast te bieden bij het reflecteren en vragen om feedback.	0.43
82 als student na te denken en in gesprek te gaan met zijn begeleider over zijn leren op een betekenisvol moment.	0.36

Aantal	SD	Variantie	Min.	Max.	Gem.	Mediaan
8	0.08	0.006	0.15	0.43	0.32	0.33

Cluster 3: Hulpmiddelen	0.17
9 modeling examples te gebruiken om de student te laten zien wat van ze verwacht wordt.	0.17
43 te voorkomen dat studenten zich blind staren op wat we stellen dat 'het proces' is.	0.2
46 de student de juiste woordenschat aan te leren.	0.17
81 het portfolio als basis te nemen waar de student zijn ontwikkelpunten op bij stelt naar gelang deze al aansluit bij het KD, ambities en dromen.	0.16

Aantal	SD	Variantie	Min.	Max.	Gem.	Mediaan
4	0.02	0	0.16	0.20	0.17	0.17

Cluster 4: Zelfsturing	0.09
6 dat de student leert zijn eigen prestatie te beoordelen op basis van zelfreflectie en externe feedback.	0.02
11 een actieve rol en (toenemende mate van) eigenaarschap van de student in het gehele beoordelingsproces door middel van scaffolding te realiseren.	0.21
12 te werken aan self-efficacy	0.3
22 studenten direct zelfstandig te leren werken.	0.04
24 studenten te leren hoe je opbouwend communiceert.	0.09
36 studenten te leren hoe zij het beste leren binnen welke context.	0.12
47 de student te leren eigen leerdoelen en leerweg te bepalen binnen de kaders.	0.03
55 de huidige student een andere beroepshouding aan te leren waarbij zij niet alleen komen 'halen' maar zelf een actievere leerhouding aannemen.	0
56 studenten inzicht te geven in hun eigen professionele identiteit.	0.07
57 studenten te leren reflecteren en formuleren van eigen leerdoelen.	0.09
62 studenten te leren hoe ze van feedback, beoordelingen en reflecties kunnen leren.	0.1
64 de student echt eigenaarschap te geven over zijn leerproces, zowel voor de inhoud als de uitvoering.	0.16
66 studenten te leren kritisch te zijn op eigen werk: zelfreflectie.	0.06
68 de student zich eigenaar en verantwoordelijk te laten voelen voor de 'beoordeling'.	0.09
71 te leren kritische en constructieve feedback te geven en te leren daarom te vragen.	0.09
74 zicht te krijgen in eigen leerproces/leerstijl.	0.09
80 de student te helpen bij het nadenken over haalbare leerdoelen.	0.06
84 dat de student de bedoeling begrijpt van een duurzame "beoordeling" voor zijn persoonlijk leerproces.	0.05

Aantal	SD	Variantie	Min.	Max.	Gem.	Mediaan
18	0.07	0.005	0	0.3	0.09	0.09

Cluster 5: Organisatiecultuur	0.25
3 formeel en non-formeel leren te erkennen.	0.24
4 de mate van duurzaamheid van kennis te erkennen.	0.23
13 dat zowel de student als docent het ingeleverde werk beoordelen.	0.24
14 tools paraat te hebben die bij onze mbo-studenten passen om het leerproces te bespreken, inzichtelijk te maken, te visualiseren en te plannen.	0.22
18 een andere term dan "beoordeling" te gebruiken.	0.26
27 in te zien hoe moeilijk reflectie en zelfbeoordeling is (de student is een mbo student) en dus is er opbouw (scaffolding) nodig.	0.25
28 de tijd te krijgen en te nemen om samen (als team) te leren en onderzoeken.	0.27
31 de begeleiding daar op af te stemmen.	0.26
34 toewijding te hebben op alle niveaus (docent --> leiding --> CvB) en het echt als SintLucas omarmen en uitdragen.	0.25
39 een duidelijke visie te hebben op beoordelen.	0.23
42 collega's de noodzaak hiervan te laten inzien.	0.26
50 te beoordelen en reflecteren met een team.	0.33

Aantal	SD	Variantie	Min.	Max.	Gem.	Mediaan
12	0.03	0.001	0.22	0.33	0.25	0.25

Cluster 6: Voorwaarden voor implementatie	0.40
15 taal paraat te hebben die bij onze mbo-studenten past om het leerproces te bespreken, inzichtelijk te maken, te visualiseren en te plannen.	0.47
25 het (duurzaam beoordelen) geïntegreerd aan te bieden als logisch onderdeel van de "practice"/"het creatieve proces".	0.29
26 er alert op te zijn "het" (duurzaam beoordelen) niet te theoretisch of analytisch aan te pakken.	0.49

29	de tijd te krijgen en te nemen om samen (als team) te bouwen aan een werkbare methodiek.	0.28
30	te blijven ontwikkelen na invoering.	0.34
35	eerlijk en kritisch plannen te maken, rekening houdend met wat haalbaar is in duurzaam beoordelen in tijd en ruimte die we hebben.	0.37
40	ruimte te geven aan collega's om te leren hoe zij dit het beste kunnen doen.	0.32
53	te onderzoeken wat en in welke mate van een mbo-student verwacht kan worden binnen het aannemen van een actieve leerhouding/tonen van eigenaarschap.	0.57
54	formatieve toetsing een plek te geven binnen ons onderwijs.	0.34
58	de rol van de docent binnen het dagelijkse onderwijs eens goed te bekijken/onderzoeken.	0.48

Aantal	SD	Variantie	Min.	Max.	Gem.	Mediaan
10	0.1	0.009	0.29	0.6	0.4	0.36

Cluster 7: Ipsatief beoordelen 0.77

8	heldere beoordelingscriteria te geven en de student te leren deze ook zelf te formuleren.	1
60	flexibel te kunnen zijn in de mate waarin je verlangt dat de student eigenaarschap over zijn/haar ontwikkelproces neemt en de verhouding tussen keuze en vaststaand programma daarop aanpast.	0.57
65	te focussen op groei.	0.81
70	kritisch te zijn op eigen handelen en dat van andere studenten/docenten.	0.71
72	vanuit de docent gezien, te beseffen dat iedere student anders leert en in een ander tempo.	0.71
79	beoordelingspunten aan te geven die duidelijk zijn voor zowel student als docent.	0.81

Aantal	SD	Variantie	Min.	Max.	Gem.	Mediaan
6	0.13	0.017	0.57	1	0.77	0.76

Cluster 8: Groeimindset	0.45
10 alle elementen van de complexe zelfsturende beoordelingsvaardigheid van de student te stimuleren, waardoor de student een geïnformeerd oordeel ontwikkelt.	0.53
20 te reflecteren op het proces, de aanpak, de context en allerlei verschillende factoren die bepalen in hoeverre iets goed of minder goed gelukt is.	0.28
21 te reflecteren voorafgaand aan beoordelen.	0.38
32 bijvoorbeeld de feedbackcyclus in te zetten waarmee goed afgestemd kan worden op de behoefte van de student op ieder moment in het leerproces/project.	0.28
52 studenten zelf het eindniveau laten bepalen.	0.25
59 een goede "mindset" aan het begin van de opleiding te bewerkstelligen (ook bij docenten); eigenaarschap over leren als logische competentie die je in de praktijk nodig hebt.	0.41
73 te leren van 'fouten', accepteren dat in een leerproces een 'fout' je verder kan helpen en als gevolg daarvan durven experimenteren.	0.43
75 terug te kijken en vooruit te kijken; hoe ging ik een jaar geleden met deze materie om? hoe doe ik dat nu? en wat zou ik willen bereiken volgend jaar?	0.71
76 beoordelen te zien als waarderen van jouw inspanning voor jezelf en voor anderen (de docent, de klant, etc.).	0.73
77 bewust te zijn/worden (student en docent) van het leerproces, wat is van belang in deze fase van je leerproces.	0.59
78 een student onder begeleiding van een docent te helpen groeien in het kunnen reflecteren, het nemen van verantwoordelijkheid en in het vooruitkijken naar de (nabije) toekomst.	0.39
83 niet uit te gaan van een vaste route. De student zal naar aanleiding van ervaringen (vakinhoudelijk, maatschappelijk, persoonlijk) steeds weer zijn leerdoelen bijstellen.	0.45

Aantal	SD	Variantie	Min.	Max.	Gem.	Mediaan
12	0.15	0.023	0.25	0.73	0.45	0.42

Bijlage D

Gemiddelde scores per statement op belang en haalbaarheid

Focus prompt: Voor het realiseren van een duurzame beoordelingspraktijk in het dagelijks onderwijs waarmee we de student helpen voorbereiden op een leven lang leren is het van belang om.....

ID	Stellingen	Gemiddelde score belang	Gemiddelde score haalbaarheid
2	de unieke student centraal te stellen door middel van een ipsatief beoordelingsmodel (groei ten opzichte van jezelf i.p.v. anderen)	4.1818	3.8182
3	formeel en non-formeel leren te erkennen.	3.8182	4.1818
4	de mate van duurzaamheid van kennis te erkennen.	3.6364	4.0000
5	de student te betrekken bij het ontwerpen van beoordelingstaken.	3.5455	3.9091
9	modeling examples te gebruiken om de student te laten zien wat van ze verwacht wordt.	3.5455	3.7273
10	alle elementen van de complexe zelfsturende beoordelingsvaardigheid van de student te stimuleren, waardoor de student een geïnformeerd oordeel ontwikkelt.	3.9091	3.7273
11	een actieve rol en (toenemende mate van) eigenaarschap van de student in het gehele beoordelingsproces door middel van scaffolding te realiseren.	4.0000	4.1818
12	te werken aan self-efficacy	4.1000	4.0909
18	een andere term dan "beoordeling" te gebruiken.	3.4545	3.8182
19	beoordelen te zien als een ontdekkingsreis voor de student, samen met zijn begeleider waarin er geen goed of fout is.	4.2727	4.0909
22	studenten direct zelfstandig te leren werken.	3.1818	3.6364
29	de tijd te krijgen en te nemen om samen (als team) te bouwen aan een werkbare methodiek.	4.2727	3.2727
32	bijvoorbeeld de feedbackcyclus in te zetten waarmee goed afgestemd kan worden op de behoefte van de student op ieder moment in het leerproces/project.	3.8182	4.0000

34	toewijding te hebben op alle niveaus (docent --> leiding --> CvB) en het echt als SintLucas omarmen en uitdragen.	3.9091	3.5455
36	studenten te leren hoe zij het beste leren binnen welke context.	3.9091	3.9091
37	studenten met verschillende typen portfolio's te laten werken.	3.8182	3.8182
43	te voorkomen dat studenten zich blind staren op wat we stellen dat 'het proces' is.	3.4545	3.8182
46	de student de juiste woordenschat aan te leren.	4.0909	4.0000
50	te beoordelen en reflecteren met een team.	4.0909	4.0909
52	studenten zelf het eindniveau laten bepalen.	3.6364	2.9091
76	beoordelen te zien als waarderen van jouw inspanning voor jezelf en voor anderen (de docent, de klant, etc.).	3.7273	3.6364
77	bewust te zijn/worden (student en docent) van het leerproces, wat is van belang in deze fase van je leerproces.	4.2727	4.1818
81	het portfolio als basis te nemen waar de student zijn ontwikkelpunten op bij stelt naar gelang deze al aansluit bij het KD, ambities en dromen.	3.8182	4.0909
1	authentieke en betekenisvolle beoordelingssituaties te realiseren.	4.3636	4.1818
15	taal paraat te hebben die bij onze mbo-studenten past om het leerproces te bespreken, inzichtelijk te maken, te visualiseren en te plannen.	4.3636	4.1818
20	te reflecteren op het proces, de aanpak, de context en allerlei verschillende factoren die bepalen in hoeverre iets goed of minder goed gelukt is.	4.5455	4.1818
28	de tijd te krijgen en te nemen om samen (als team) te leren en onderzoeken.	4.3636	3.5455
40	ruimte te geven aan collega's om te leren hoe zij dit het beste kunnen doen.	4.3636	4.0909
42	collega's de noodzaak hiervan te laten inzien.	4.6364	4.0000
47	de student te leren eigen leerdoelen en leerweg te bepalen binnen de kaders.	4.3636	3.8182
48	studenten ruimte/veiligheid te bieden voor het maken van fouten en verbeteren (na de beoordeling) zonder vertraging op te lopen.	4.6364	4.1818

55	de huidige student een andere beroepshouding aan te leren waarbij zij niet alleen komen 'halen' maar zelf een actievere leerhouding aannemen.	4.7273	4.1818
56	studenten inzicht te geven in hun eigen professionele identiteit.	4.3636	4.1818
59	een goede "mindset" aan het begin van de opleiding te bewerkstelligen (ook bij docenten); eigenaarschap over leren als logische competentie die je in de praktijk nodig hebt.	4.3636	4.1818
60	flexibel te kunnen zijn in de mate waarin je verlangt dat de student eigenaarschap over zijn/haar ontwikkelproces neemt en de verhouding tussen keuze en vaststaand programma daarop aanpast.	4.3636	4.1818
61	de student te stimuleren eerlijk en onbevangen naar zichzelf en zijn omgeving te kijken.	4.5455	4.0909
64	de student echt eigenaarschap te geven over zijn leerproces, zowel voor de inhoud als de uitvoering.	4.4545	3.8182
68	de student zich eigenaar en verantwoordelijk te laten voelen voor de 'beoordeling'.	4.5455	4.1818
70	kritisch te zijn op eigen handelen en dat van andere studenten/docenten.	4.6364	4.0000
71	te leren kritische en constructieve feedback te geven en te leren daarom te vragen.	4.6364	4.1818
82	als student na te denken en in gesprek te gaan met zijn begeleider over zijn leren op een betekenisvol moment.	4.3636	4.0000
83	niet uit te gaan van een vaste route. De student zal naar aanleiding van ervaringen (vakinhoudelijk, maatschappelijk, persoonlijk) steeds weer zijn leerdoelen bijstellen.	4.5455	3.7273
8	heldere beoordelingscriteria te geven en de student te leren deze ook zelf te formuleren.	4.2727	4.5455
17	het voor de student uitdagend te maken.	4.2727	4.4545
23	studenten te leren reflecteren op elkaars werk.	4.2727	4.2727
25	het (duurzaam beoordelen) geïntegreerd aan te bieden als logisch onderdeel van de "practice"/"het creatieve proces".	4.2727	4.2727
26	er alert op te zijn "het" (duurzaam beoordelen) niet te theoretisch of analytisch aan te pakken.	4.0000	4.2727

33	studenten te bevragen hoe zij dit voor zich zien.	4.1818	4.4545
35	eerlijk en kritisch plannen te maken, rekening houdend met wat haalbaar is in duurzaam beoordelen in tijd en ruimte die we hebben.	4.0909	4.3636
41	vertrouwen te hebben in de goede wil van de studenten.	4.2727	4.3636
53	te onderzoeken wat en in welke mate van een mbo-student verwacht kan worden binnen het aannemen van een actieve leerhouding/tonen van eigenaarschap.	4.0909	4.2727
75	terug te kijken en vooruit te kijken; hoe ging ik een jaar geleden met deze materie om? hoe doe ik dat nu? en wat zou ik willen bereiken volgend jaar?	4.0909	4.6364
6	dat de student leert zijn eigen prestatie te beoordelen op basis van zelfreflectie en externe feedback.	4.6364	4.4545
7	dat zowel student als docent informatierijke feedback geeft en ontvangt.	4.6364	4.3636
13	dat zowel de student als docent het ingeleverde werk beoordelen.	4.5455	4.4545
14	tools paraat te hebben die bij onze mbo-studenten passen om het leerproces te bespreken, inzichtelijk te maken, te visualiseren en te plannen.	4.3636	4.4545
16	het voor de student praktisch te maken.	4.6364	4.5455
21	te reflecteren voorafgaand aan beoordelen.	4.5455	4.4545
24	studenten te leren hoe je opbouwend communiceert.	4.3636	4.2727
27	in te zien hoe moeilijk reflectie en zelfbeoordeling is (de student is een mbo student) en dus is er opbouw (scaffolding) nodig.	4.3636	4.2727
30	te blijven ontwikkelen na invoering.	4.3636	4.2727
31	de begeleiding daar op af te stemmen.	4.5455	4.4545
38	leren centraal te stellen.	4.5455	4.5455
39	een duidelijke visie te hebben op beoordelen.	4.4545	4.2727
44	de student zich veilig te laten voelen.	4.9091	4.9091
45	de student kennis te laten maken met opbouwende feedback.	4.6364	4.6364
49	tijd te nemen voor overleg en het gesprek aangaan met de student.	4.8182	4.2727
51	de student te kennen.	4.6364	4.4545

54	formatieve toetsing een plek te geven binnen ons onderwijs.	4.4545	4.2727
57	studenten te leren reflecteren en formuleren van eigen leerdoelen.	4.5455	4.4545
58	de rol van de docent binnen het dagelijkse onderwijs eens goed te bekijken/onderzoeken.	4.5455	4.2727
62	studenten te leren hoe ze van feedback, beoordelingen en reflecties kunnen leren.	4.7273	4.5455
63	studenten houvast te bieden bij het reflecteren en vragen om feedback.	4.6364	4.6364
65	te focussen op groei.	4.3636	4.6364
66	studenten te leren kritisch te zijn op eigen werk: zelfreflectie.	4.7273	4.2727
67	nieuwsgierigheid te prikkelen.	4.3636	4.3636
69	de student niet af te rekenen op wat niet goed gaat, maar te letten op wat goed ging en wat beter kan.	4.6364	4.5455
72	vanuit de docent gezien, te beseffen dat iedere student anders leert en in een ander tempo.	4.6364	4.5455
73	te leren van 'fouten', accepteren dat in een leerproces een 'fout' je verder kan helpen en als gevolg daarvan durven experimenteren.	4.8182	4.5455
74	zicht te krijgen in eigen leerproces/leerstijl.	4.7273	4.2727
78	een student onder begeleiding van een docent te helpen groeien in het kunnen reflecteren, het nemen van verantwoordelijkheid en in het vooruitkijken naar de (nabije) toekomst.	4.4545	4.3636
79	beoordelingspunten aan te geven die duidelijk zijn voor zowel student als docent.	4.4545	4.6000
80	de student te helpen bij het nadenken over haalbare leerdoelen.	4.4545	4.7273
84	dat de student de bedoeling begrijpt van een duurzame "beoordeling" voor zijn persoonlijk leerproces.	4.3636	4.2727